Soil Biology and Biochemistry 127 (2018) 264–279

Contents lists available at ScienceDirect

Soil Biology and Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Environmental drivers of the geographical distribution of methanotrophs: Insights from a national survey

Loïc Nazaries[a,*], Senani B. Karunaratne[a], Manuel Delgado-Baquerizo[a,b], Colin D. Campbell[c,d], Brajesh K. Singh[a,e,**]

[a] Hawkesbury Institute for the Environment, Western Sydney University, Building L9, Locked Bag 1797, Penrith, NSW, 2751, Australia
[b] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, 80309, USA
[c] The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8BQ, Scotland, UK
[d] Department Soil and Environment, Swedish Agricultural Sciences University, Uppsala, Sweden
[e] Global Centre for Land-Based Innovation, Western Sydney University, Locked Bag 1797, Penrith, NSW, 2751, Australia

ARTICLE INFO

Keywords:
Methanotrophs
pmoA
Spatial modelling
Mapping
Niche partitioning
Biogeography

ABSTRACT

There is considerable evidence that environmental properties are important for microbial niche partitioning in general. However, little is known about the environmental factors explaining this for soil methane-oxidising bacteria (or methanotrophs), which play an essential role in ecosystem functioning and climate regulation through mitigation of net CH_4 emissions worldwide. This knowledge gap limits the inclusion of taxon-based information to improve predictions of climate change-simulation models. In this study, 697 soil samples were collected across Scotland and 62 climo-edaphic properties were analysed. Combined with a set of hybrid geostatistical modelling approaches, the aim of this study was to investigate the biogeographical distribution (*pmoA* gene relative abundance) of key methanotrophic operational taxonomic units named Terminal-Restriction Fragments (T-RFs) and of methanotrophic community structure. The main objectives were to: 1) identify major environmental drivers influencing the distribution and composition of methanotrophs; and 2) perform spatial modelling and mapping of soil methanotrophic community assemblage and distribution of those dominant T-RFs. Herein, it was hypothesised that the assemblage of methanotrophic community and distribution of key populations across various landscapes could be predicted using a range of climo-edaphic factors optimised for spatial, climate and terrain attributes. The findings presented here suggest that the distribution of methanotrophs is strongly linked to land use and some edaphic properties, predominantly soil moisture/rainfall, nutrients and metal ions. The hybrid geostatistical approach allowed for spatial prediction of methanotrophic T-RFs and community, and demonstrated a clear niche partitioning between dominant T-RFs. Overall, these results provide novel evidence that the distribution of methanotrophs could be explained and mapped in terms of niche partitioning and predicted at the regional scale. The findings of the present study have significance for the sustainable management of ecosystems and improvement of simulation models for better prediction of ecosystem functions under predicted global changes.

1. Introduction

Planet Earth relies on the presence, diversity and abundance of soil microorganisms, and as such, are critical for the maintenance and stability of ecosystem functions, including primary productivity and climate regulation (Delgado-Baquerizo et al., 2016b; Singh et al., 2010). Identifying the environmental drivers and processes which influence the distribution of microbial taxa is important in order to understand the fundamental ecological and evolutionary processes but also to predict ecosystem response to global change (Singh et al., 2010). Climate, soil organic matter, spatial distance and soil pH are currently considered to be critical drivers of microbial community at different levels spanning from taxonomic level to various geographical scales (Fierer and Jackson, 2006; Horner-Devine et al., 2004). It was consequently proposed that ecological theories developed for plant and animals could predict microbial distribution, including niche partitioning (Lennon et al., 2012). Within the scope of the current paper, niche partitioning (local/global process-based environments such as

* Corresponding author.
** Corresponding author. Hawkesbury Institute for the Environment, Western Sydney University, Building L9, Locked Bag 1797, Penrith, NSW, 2751, Australia.
E-mail addresses: loicnazaries@yahoo.fr (L. Nazaries), B.Singh@westernsydney.edu.au (B.K. Singh).

https://doi.org/10.1016/j.soilbio.2018.08.014
Received 3 October 2017; Received in revised form 30 July 2018; Accepted 17 August 2018
Available online 19 September 2018
0038-0717/

methanotrophy) may represent a space wherein local bio-edaphic properties drive microbial abundance, diversity and richness (methanotrophs) pending nutrient accessibility and/or affinity (methane) (Knief, 2015; Krause et al., 2014; Pratscher et al., 2018).

Some studies on microbial niche distribution challenge this unifying concept and suggest that drivers shaping the microbial community assemblage may vary depending on the geographical scale of study (*e.g.* global or regional) (Delgado-Baquerizo et al., 2016c; Powell et al., 2015a; Ranjard et al., 2010). Much less is known on the environmental drivers of specialised microbial communities and large-scale studies addressing this important knowledge gap are lacking (Powell et al., 2015a; Yao et al., 2013). Specialised soil microbial communities such as methanotrophs (but also, nitrifiers, nitrogen-fixing bacteria, denitrifiers) represent excellent candidates for testing ecological distribution models due to their lesser diversity and complexity compared to whole microbial community. According to recent publications (Dumont et al., 2014; Lüke et al., 2010; Wen et al., 2016), the cut-off values for sequence identity of the *pmoA* gene may be altered compared to 16S rRNA gene with the following: 86% sequence identity at species level (*vs.* 97% for 16S rRNA); 83% for genus level (*vs.* 95% for 16S rRNA); and 71% for family level (*vs.* 90% for 16S rRNA). This new cut-off value for phylogentic assignment permits a more precise estimation of methanotroph variability in soils (Wen et al., 2016).

Methanotrophs play an important role in the global methane (CH_4) budget. Both the production and consumption of CH_4 are microbiologically-driven processes (Conrad, 2009, 2007). The main microbial process for CH_4 production, called methanogenesis, occurs in anaerobic conditions (*e.g.* in rice paddies, landfills, termites, ruminants, *etc.*) through the activity of methanogens mainly belonging to *Archaea* (Nazaries et al., 2013a). On the other hand, the oxidation of CH_4 by microbes (or methanotrophy) mainly takes place at the aerobic and oxic/anoxic interface of various habitats. Recently, it was shown that there exist alternative processes of methane oxidation in anaerobic conditions (sometimes known as reversed methanogenesis) (Thauer and Shima, 2008). Other newly proposed paths involved the implication of nitrite and sulphate as electron accepter when in direct contact with surrounding bacteria (Ettwig et al., 2010; Knittel and Boetius, 2009). The present study only focused on aerobic methanotrophs. Methanotrophs are especially important in terrestrial ecosystems as they constitute the main biological control over the release of CH_4 from the deeper soil layers to the atmosphere (such as in wetlands) and thus play a major role in mitigating CH_4 emissions world-wide. Past studies suggested that methanotrophic populations, dominated by low-affinity methanotrophs (*e.g.* α-proteobacterial methanotrophs such as *Methylocystaceae*, *Methylococcaceae* or γ-proteobacterial methanotrophs such as *Methylococcus* spp.), are mainly found in CH_4-emitting environments, while high-affinity methanotrophs (*e.g.* α-proteobacterial methanotrophs such as USCα or γ-proteobacterial methanotrophs such as USCγ) thrive better in CH_4-sink ecosystems (Knief, 2015). There are alternatives as well, because *Methylocystis* members were detected to contribute to atmospheric CH_4 sink, in particular in ecosystems such as deciduous forests and/or displaying soil attributes such as acidic-to-neutral pH (Kolb, 2009). Overall, high-affinity methanotrophs constitute the only natural biological sink of atmospheric CH_4 into soils, estimated to account for about 5% of total removal (Reay et al., 2010). Consequently, improving our knowledge of the ecological drivers of methanotrophs is critical to predict the environmental change feedbacks from local to global scale. This implies that a better understanding of the distribution patterns of these organisms is needed to unravel the biological mechanisms of CH_4 source/sink and find ways to mitigate global CH_4 emissions.

Mapping microbial communities and important taxa can provide an insight into ecological processes at diverse geographical scales (*i.e.* regional/national/global). Mapping the geographical (either at small or large scale) distribution of certain micro-organisms can help formulating policies and management decisions for better sustainability outcomes. While several studies attempted to map terrestrial microbial communities at a regional scale with a range of experimental and statistical approaches (Griffiths et al., 2011; Hermans et al., 2017; Larsen et al., 2012), the investigators' efforts focused on the detection of 16S rRNA genes using universal bacterial primer sets. Additionally, other studies concentrated on the regional drivers of the spatial distribution of soil microbial communities, especially for bacteria in the broad sense but also for specific ecosystem functions such as nitrification/denitrification (*e.g.* Bru et al., 2011; Dequiedt et al., 2011; Constancias et al., 2015). Only a very limited number of studies (Krause et al., 2013, 2009; Lüke et al., 2010) attempted to investigate the spatial patterns of methanotrophs but this was only at the scale of the plot (m^2 scale) and for a very specific habitat (wetlands comprising riparian floodplain or rice paddies) and therefore, had a limited scope to clearly identify drivers of methanotrophic spatial distribution on a much larger scale (hectare scale). Nonetheless, previous studies suggest that methanotrophic community composition is driven, at the plot-scale, by a wide variety of environmental (climo-edaphic) properties such as soil pH, temperature, nutrient and water content (rainfall), CH_4/O_2 gas gradient and soil texture at small-to-medium scales (Kolb, 2009; Tate, 2015).

Because both simulation models and policy decisions require information at a larger scale, the need for a better mechanistic understanding of the large-scale drivers of specialised microbes – such as methanotrophs – has emerged. Additionally, mapping the methanotrophic distribution can improve the prediction of their response to future environmental changes. To achieve this, the following knowledge gaps need to be addressed: 1) identification of important drivers of methanotroph distribution; 2) decipherment of how those environmental drivers impact on methanotrophic community distribution; 3) validation of a link between diversity and function (*i.e.* is there a specific relationship between methanotroph distribution patterns and the ability of an ecosystem to mitigated CH_4 emissions?) at various geographical scales; and 4) integration of microbial parameterisation, such as the inclusion of community composition data, in order to reduce the uncertainties of predictive models. Overall, achieving these milestones could give the researchers the necessary tools to, possibly, manage ecosystems in such a way that soils could potentially become sink for CH_4.

In an attempt to elucidate some of the knowledge gaps enumerated above, the present study had two main objectives: 1) identify the key environmental drivers influencing the distribution of the most important and abundant methanotrophic taxa using statistical modelling; and 2) perform spatial modelling and mapping the soil methanotrophic community assemblage as well as the distribution of those dominant taxa. Herein, it was hypothesised that the assemblage of the methanotrophic community and the distribution of key population across various landscapes could be predicted using a range of climo-edaphic factors optimised for spatial, climate and terrain attributes. Numerous bioclimatic models for the prediction of biological assemblages exist (see references in Larsen et al., 2012), but their use has primarily been for the study of macroecology and seldom in the field of microbial ecology (Larsen et al., 2012). We posit that a similar approach of modelling could be used to identify the main drivers of methanotrophic communities, as well as their distribution at a regional scale.

2. Methods

2.1. Study area and dataset

Here, the National Soils Inventory of Scotland re-sampling (NSIS_2) dataset was explored, as described by Lilly et al. (2010a, 2010b). Briefly, an extensive soil survey, following a $20 \times 20\,km^2$ sampling grid, was conducted across Scotland ($77{,}933\,km^2$), a Northern, temperate boreal climate country, between 2006 and 2009. This dataset includes 183 sites which vegetation was aggregated into six broad land use types: arable land, bog, moorland, semi-natural grassland,

improved grassland and woodland. For each of the 183 sites, four replicates were sampled 4, 8, 16 and 32 m from a central pit. A few sites included three replicates only due to the nature of the terrain (Yao et al., 2013). In total, 713 samples were analysed for a wide range of climo-edaphic properties: latitude/longitude; altitude; terrain slope; soil drainage; mean annual precipitation and temperature; maximum root depth; NH_x and NO_x deposition; N fertiliser input; and 51 soil properties (pH, Total C, N and P, soil C/N ratio, field moisture, NH_4^+ and NO_3^- concentrations, bulk density; sand, silt, clay; oxalate extractable Al, Fe, Mn, P and Si; Aqua Regia (AQR) extracted elements, specifically, Ag, As, Ba, Cd, Co, Cr, Cu, Hg, Mo, Ni, Pb, Pt, Se, Sr, Zn, Al, B, Ca, Fe, K, Mg, Mn, Na, P, S and Ti; exchangeable Ca, Mg, Na, K, Mn, Fe, Al and H). Note that "soil drainage" represents the removal of excess water and thus, is closely related to soil moisture, bulk density and sand/silt/clay content. Other soil characteristics related to soil formation (*e.g.* soil parental material and related rock forming characteristics) were not all available. These variables refer to soil forming factors (Jenny, 1994) and would have certainly enhanced the quality of the models tested. More details about the NSIS_2 dataset, including site location and protocols for soil sampling and analysis can be found elsewhere (Lilly et al., 2010a, 2010b; Yao et al., 2013). Overall, this dataset (see Appendix S 1) provides a vast gradient of 62 climo-edaphic properties. Shortly after collection (a few hours), soil samples were brought back to the laboratory and kept in a 4 °C cold room until physico-chemical analyses were preformed (within a week). In parallel, a subset of the soil samples (~5 g) was stored at −80 °C upon arrival at the laboratory. DNA extraction (see below) was then carried out within a few weeks.

2.2. Analysis of the methanotrophic community assemblage

Terminal-restriction fragment length polymorphism (T-RFLP) analysis was chosen for resolving gene sequence diversity of the particulate methane monooxygenase (pMMO) enzyme which constitutes the key enzyme responsible for the oxidation of CH_4 and is found in all methanotrophs, except in *Methylocella* and *Methyloferula* spp. (see references in Knief, 2015). Despite a lower resolution compared to more recent deep-sequencing approaches, T-RFLP is especially suitable (Delgado-Baquerizo et al., 2016a) for determining microbial community composition. Specialised microbial groups (such as methanotrophs) can be investigated using functional genes wherein the diversity is low and these groups represent only a minor fraction of the overall microbial community (Hu et al., 2015).

Following soil DNA extraction (see Yao et al., 2013 for details), PCR and T-RFLP analyses of the *pmoA* genes were carried out (Nazaries et al., 2013b; Yao et al., 2013). Briefly, total DNA was extracted from 0.5 g of soil and the primer set pmo189F/pmo650R (Bourne et al., 2001), with the 6-FAM fluorescent label attached to the forward primer's 5′-end, was used to amplify the *pmoA* gene pool. The PCR products were purified and digested with the restriction enzyme *Hha*I. The resulting digests were separated by electrophoresis and their "size" (or peak height – relative fluorescent unit – rfu) was detected on an ABI Prism® 3030*xl* genetic analyser (Applied Biosystems, UK). A fragment size was estimated from an internal molecular ladder LIZ-labelled GeneScan™-500 internal size standard (Applied Biosystems, UK) by converting rfu into base-pair (bp) unit. Each peak relative abundance was calculated from the fluorescence of a peak against the total fluorescence of an electropherogram (for each samples). During the analysis of the T-RFLP data, 16 samples (out of 713 – see above) were removed from the dataset due to insufficient quality of the restriction profiles. Thus, the final dataset contained 697 samples (Appendix S 1). A workflow diagram of the (six) main steps (or "segments") of the analytical procedure is presented in Fig. 1.

2.3. Preparation of the T-RFLP dataset for statistical analysis

T-RFLP methodology to characterise methanotrophs is well-established and described (Nazaries et al., 2015, 2013b; Singh et al., 2009). Briefly, T-RFLP peaks which relative abundance was less than 0.01% were converted to zeros. Then, for each T-RFLP peak, the number of zeros was calculated and expressed as a percentage of the total number of samples. There may be inherent, small variations in peak's length/position, meaning that the peaks need to be "binned". This was done using a two-base pair window (Culman et al., 2009, 2008) with the resulting bins named using the size of the most dominant peak. These groups of bins/classes were then referred to as terminal-restriction fragment (T-RF). Lastly, the bins which contained more than 95% of zeros across the dataset were removed from the analysis. The subsequent dataset contained a total of 58 T-RFs (out of 59; *i.e.* the non-specific T-RF *pmoA*-499 was not included in the dataset), and was used for both community structure and T-RFs relative abundance analyses (see dark grey segment #1 in Fig. 1). The last step of cloning/sequencing was not performed in this study for experimental reasons but mainly because of time and labour involved as the aim of this paper was to study the biography of dominant methanotrophic T-RFs.

Finally, the reduced T-RFLP data matrix was standardised using the Hellinger transformation (Legendre and Gallagher, 2001). Principal component analysis (PCA) was performed on the transformed data and sample scores were extracted for the first three principal components (PCs) with the purpose of mapping methanotrophic community assemblage. The reader is referred to the Supplementary Material section for further details on PCA. The analysis was carried out using the "vegan" package (Oksanen et al., 2013) in R statistical programming language (https://cran.r-project.org/). Results of the PCA, including a bi-plot and scree plot, are summarised in Figure S 1 of Appendix S 3.

2.4. Identification of key environmental predictors and their influence on the main methanotrophic T-RFs

For this study, the three most abundant *pmoA* T-RFs were selected when considering the entire *Hha*I-generated T-RFLP dataset. These were *pmoA*-33, *pmoA*-81 and *pmoA*-130 (the numbers 33, 81 and 130 indicating the base-pair length of the T-RFs). They accounted for, respectively, 33%, 22% and 8% (63% cumulatively – see Fig. 2 and Table S 1 of Appendix S 2) of the total T-RF abundance across all the land uses investigated. To identify the main predictors of the major methanotrophic T-RFs a three-step approach was adopted (see brown segment #2 in Fig. 1), as described in great details in Appendix S 2. Briefly, Spearman's rank correlation was first used to explore the relationships between the relative abundance of the three main methanotrophic T-RFs and the spatial, climatic and edaphic properties incorporated in this study (see blue segment #3 in Fig. 1). This step was essential for pre-selecting the main factors relating to the response variables (the T-RFs *pmoA*-33, *pmoA*-81 and *pmoA*-130) due to the large number of predictor in the NSIS_2 dataset (see Appendix S 1). Secondly, using the predictors that showed to be significantly related to each of the selected methanotrophic T-RFs, the main predictors (land use; spatial, climatic and edaphic properties) of each methanotrophic T-RF were evaluated by using classification Random Forest (RF) analysis as explained in Delgado-Baquerizo et al. (2015). Finally, following identification of the most important predictors of the three main T-RFs investigated in this study, structural equation modelling (SEM) was used to evaluate the direct, indirect and total effects of the selected environmental variables (top ten most important and significant variables from the RF analyses – see Fig. 3) on each T-RF (see brown segment #2 in Fig. 1). To aid with the final interpretation of the SEMs, the standardised total effects of the environmental variables on each T-RF were also calculated. In each model, the land use (improved grassland, arable land, bog, moorland and woodland) was a categorical (dummy) variable with two levels: 1 (a land use) and 0 (the remaining land uses + semi-natural grassland).

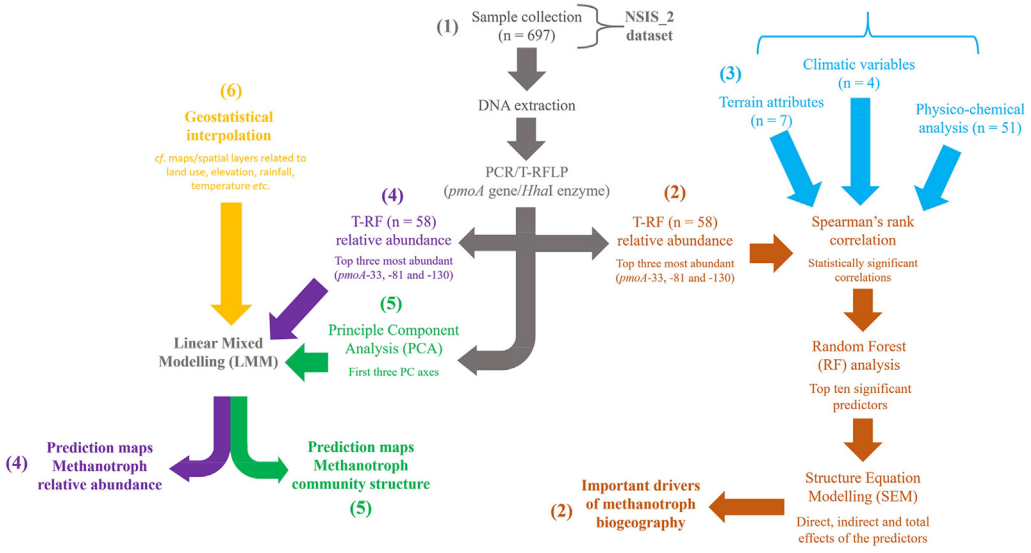

Fig. 1. Workflow diagram of the steps involved in modelling methanotroph community structure, relative abundance of the most abundant T-RFs and their environmental drivers.
Initial number of samples, environmental variables and terminal-restriction fragments (T-RFs) are given in brackets. The six "segments", or major steps, necessary for modelling the response variables are as follows: step (1) molecular analyses for isolating methanotrophic T-RFs; (2) most important drivers of the relative abundance of the three T-RFs investigated; (3) climatic and soil physico-chemical properties measured in Scotland and driving methanotrophic relative abundance and biogeography; (4) prediction maps of relative abundance of the three most abundant T-RFs; (5) prediction maps of the first three principal components of methanotrophic community structure; and (6) spatial, climate and terrain attributes available for geostatistical interpolation.

This approach allowed the comparison of the effect of a land use (*e.g.* woodland) on each microbial module in relation to the average of the remaining land uses. Note that the microsite named 'semi-natural grassland' was selected as the baseline condition (*i.e.* procedural control), and was not included explicitly in the model (Grace, 2006). When more than one variable from a specific group of variables (*e.g.* metals, minerals, organic matter) had a significant effect on a main methanotrophic T-RF, those variables were integrated as a composite variable into the model. The use of composite variables does not alter the underlying assumptions of SEM, but collapses the effects of multiple conceptually-related variables into a single composite effect, aiding interpretation of the model results (Shipley, 2016). The net (total) influence that a given variable had upon another was computed by summing all direct and indirect pathways between these two variables. Within SEM, the path coefficients designate regression coefficients as in Fig. 5. All analyses were independently done for the three T-RFs investigated.

2.5. Spatial modelling and mapping of methanotrophs

Two types of modelling/mapping were achieved. First, the relative abundance of the three most dominant T-RFs (Appendix S 2); and second, the T-RFLP profiles (Appendix S 3). The difference between the two approaches is that PCA integrated the totality of the T-RFs within the T-RFLP profiles. For instance, it can provide information on microbial diversity and/or community structure but also to test the effect of rare T-RFs. A detailed description of the approach for spatial modelling is accessible from Appendix S 2 for T-RF abundance and Appendix S 3 for community structure. To summarise, mapping the *pmoA* functional genes was completed on the relative abundance of the three selected T-RFs (see purple segment #4 in Fig. 1). Mapping was carried out using a hybrid geostatistical modelling approach by means of linear mixed-effect model (LMM) following the methodology described by Lark et al. (2006). The authors give more theoretical background and practical use of LMM in a spatial context and compared to other popular geostatistical methods such as regression-kriging. The predictors considered for spatial modelling were those that fitted two

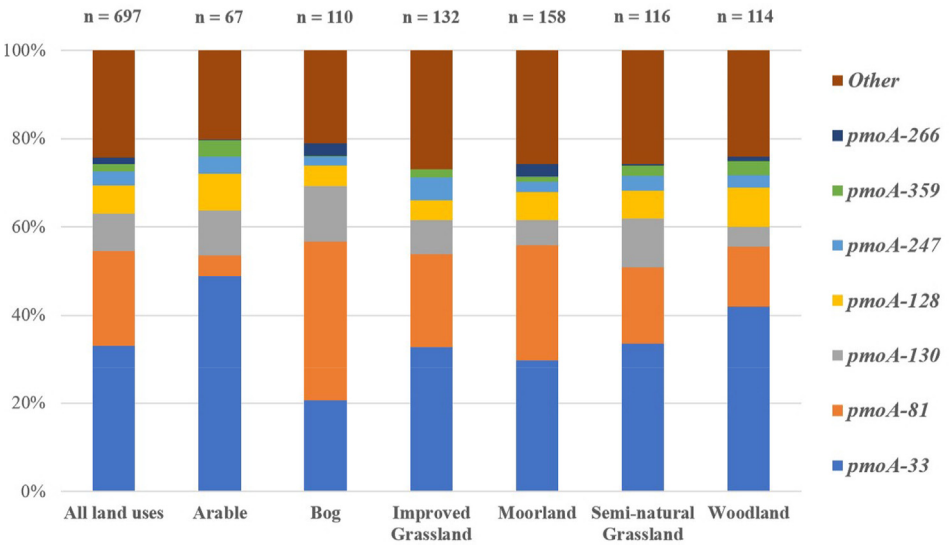

Fig. 2. Relative abundance of methanotrophic T-RFs across Scotland.
The *pmoA* genes were amplified using the primer pair pmo189F/pmo650R. The restriction enzyme for digesting *pmoA* was *Hha*I. Data represent the *Hha*I-generated T-RFs from 697 soil samples. Only T-RFs with a relative abundance above 1% are displayed, while those with abundance less than 1% were grouped in the "Other" category, which also contained the non-specific T-RF *pmoA*-499. The latter represents non-specific T-RFs (*i.e.* it does not contain restriction sites for the *Hha*I enzyme) and was detected in 11% of the samples. There was a total of 59 T-RFs (including the non-specific T-RF *pmoA*-499) in the NSIS_2 dataset. The top three T-RFs selected for this study (*pmoA*-33, *pmoA*-81 and *pmoA*-130) accounted for 63% of the total abundance and are representatives of α-proteobacterial methanotrophs. The number of samples (n) for all land uses together and for each land use are provided in bracket above each bar. The numerical (percentage) values of each category

are found in Table S 1 of Appendix S 2.

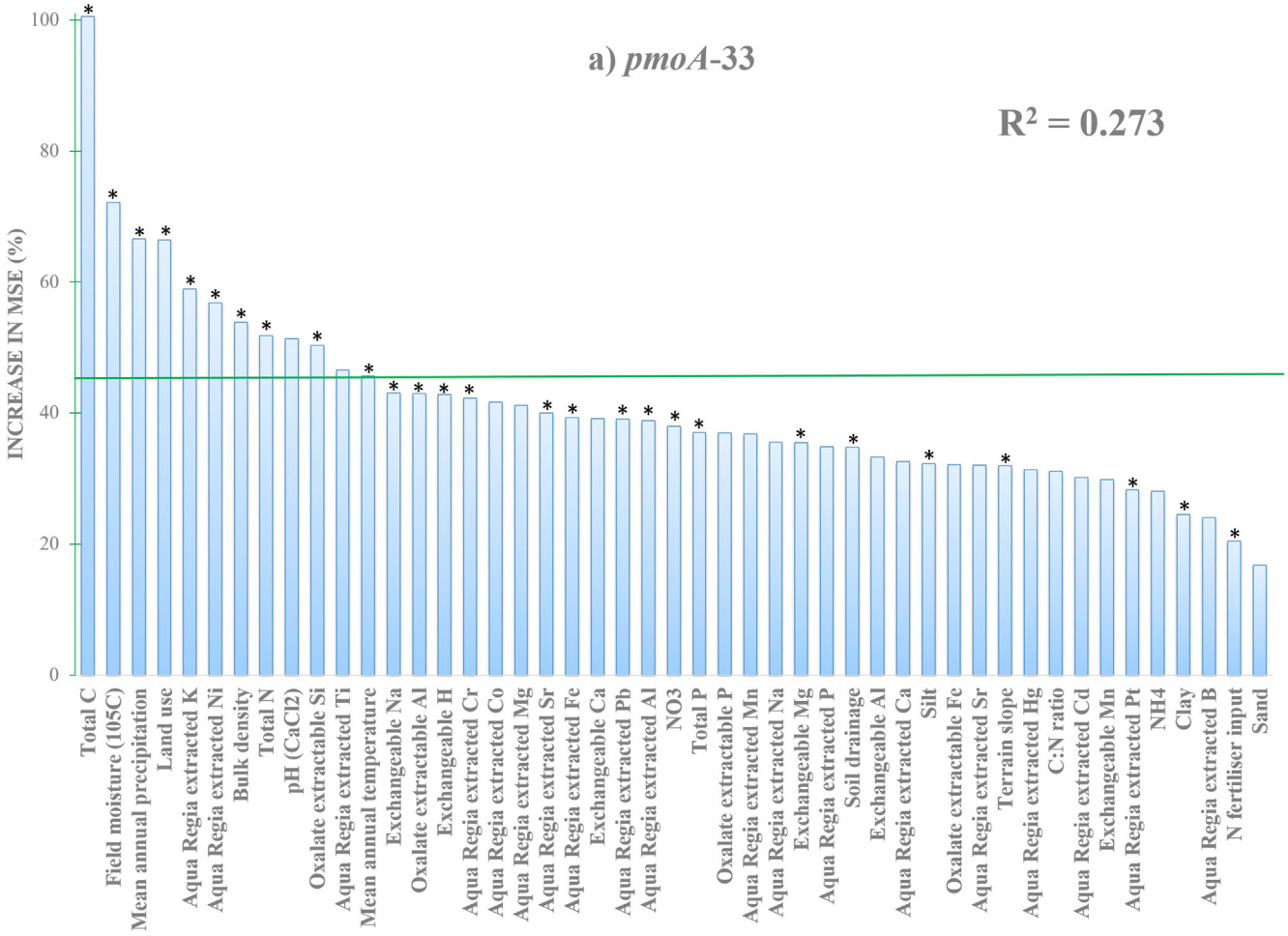

Fig. 3. Main environmental drivers of the distribution of the selected T-RFs *pmoA*-33 (a), *pmoA*-81 (b) and *pmoA*-130 (c) across Scotland.
Data represent the random forest (RF) mean predictor importance (% of increase in mean square error, MSE) of selected environmental factors. This accuracy importance measure was computed for each tree and averaged over the forest (5000 trees). Only the statistically relevant predictors (following Spearman's rank correlation tests – see Table S 2 of Appendix S 2) were included in the RF analyses. Stars above the columns indicate significant results ($P < 0.05$, see Appendix S 2). The green horizontal line indicates the threshold for the first ten statistically significant variables included in the structural equation modelling. The variables displayed represent the predictors used for LMM and geostatistical interpolation (see Fig. 5). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

parameters: 1) driver importance for *pmoA* gene distribution as identified in the RF and SEM analyses; and 2) parameter availability as spatial layers covering the study area. Other, non-available parameters were discarded. For the development of the spatial models, separate LMMs were fitted for each of the selected T-RFs (see orange segment #6 in Fig. 1). The application of LMM in geostatistical modelling is explained in more details by Lark et al. (2006) and Nelson et al. (2011). Once the optimum spatial model was developed using the observed data, predictions were carried out for un-sampled locations using the empirical best linear unbiased predictor (E-BLUP) approach, which is a general-purpose name that encapsulates the various forms of kriging (Orton et al., 2014). In order to assess the model quality, four measures were calculated as described by Karunaratne et al. (2014): mean error (ME), root mean square error (RMSE), mean standardised squared deviation ratio (MSDR) and Lin's concordance correlation coefficient (CCC). The reader is referred to Appendix S 2 for a detailed procedure for estimating the model goodness-of-fit and cross-validation.

3. Results

Knowledge of the biogeographical distribution of methanotrophs may bring critical information for future application such as modelling CH_4 mitigation strategies and which environmental properties are likely responsible. Conversely, the knowledge of methanotroph community structure (PC analysis) and biogeographical distribution will help with understanding methanotroph responses to changes in climo-edaphic properties and land-use change.

In this Results section, the focus on T-RF distribution was favoured in place of community structure (see purple segment #4 *vs*. green segment #5 in Fig. 1). For clarity, the part of the PCA results was moved to the Supplementary Material section while being briefly described within this part of the manuscript. However, a parallel between T-RF and PCA outputs was made later.

3.1. Identification of methanotrophic (pmoA) T-RFs

The three most dominant methanotrophic T-RFs identified were called *pmoA*-33, *pmoA*-81 and *pmoA*-130, which together accounted for ≈ 63% of the total relative abundance (Fig. 2 and Table S 1 of Appendix S 2). Previous work based on T-RFLP combined with cloning/sequencing and *in silico* analyses of the PubMed dataset (Nazaries et al., 2013b, 2011), identified *Hha*I-digested T-RFs (taxa) as members of the methanotrophic clades USCα (*pmoA*-33), *Methylosinus/Methylocystis* (*pmoA*-81) and Cluster 5 (*pmoA*-130) (see Table 1). Despite the lack of direct "proof", the inferences (and their strength) made here relied on the accuracy and repeatability of the T-RFLP method (Culman et al.,

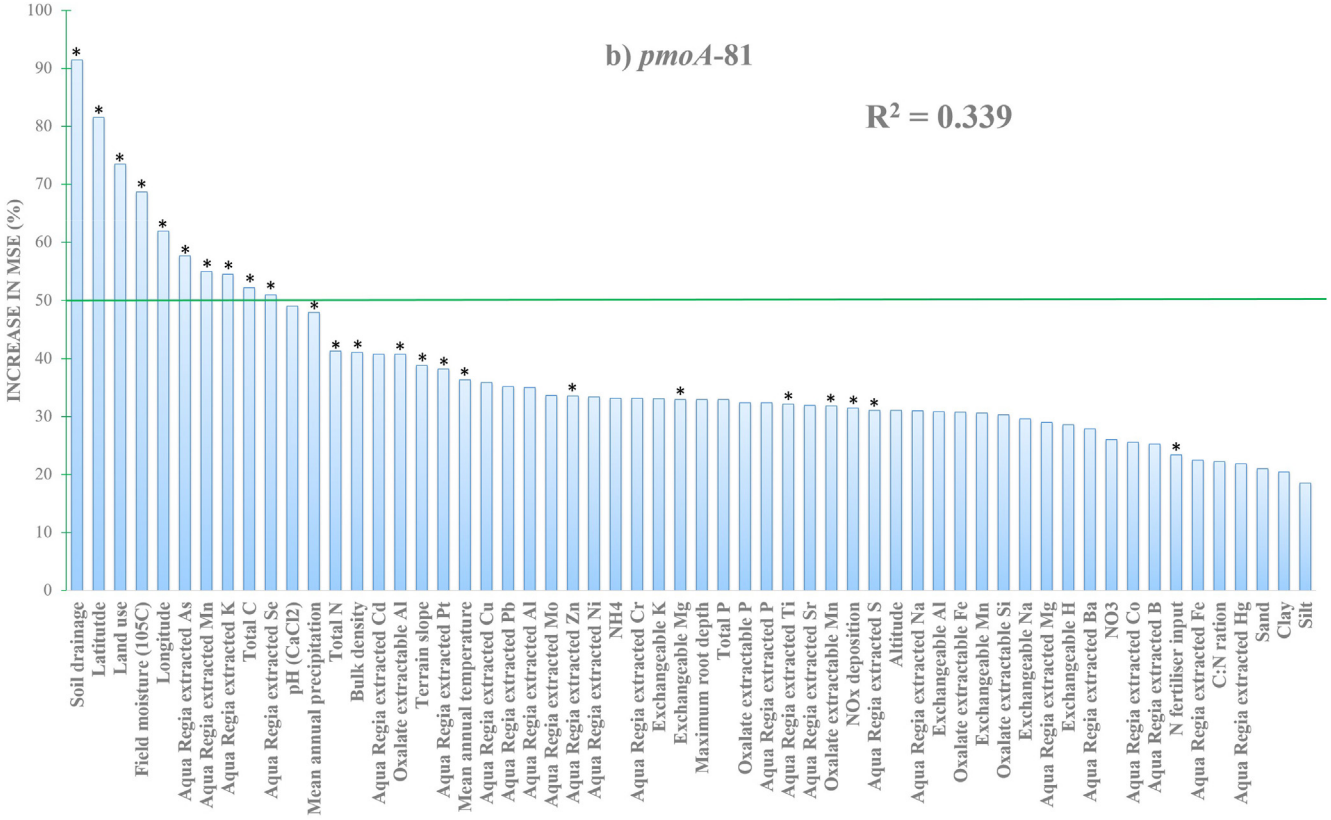

Fig. 3. (*continued*)

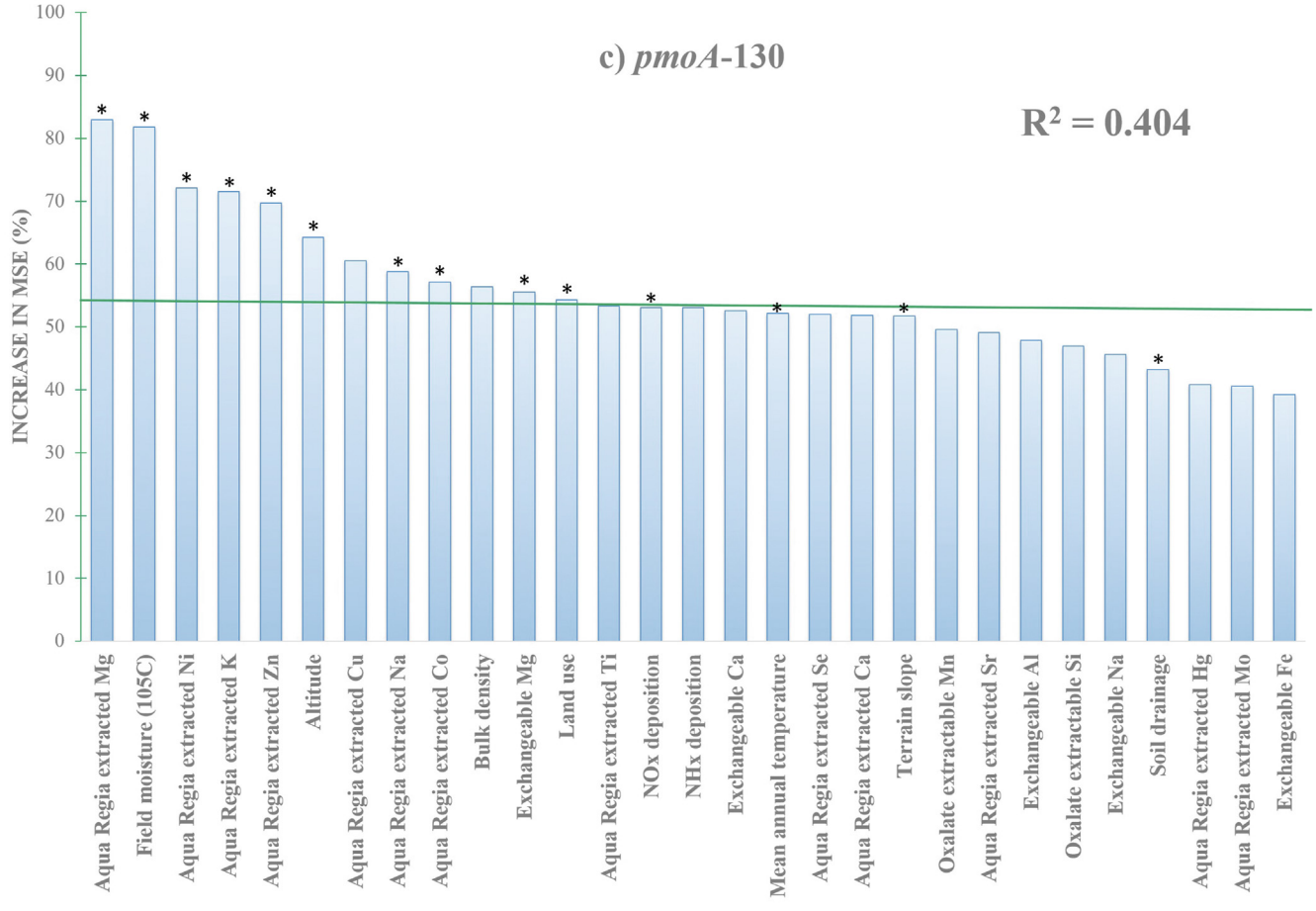

Fig. 3. (*continued*)

Table 1
Predictors of relative abundance of selected methanotrophic T-RFs in Scotland
Data below represent a summary of the statistically significant predictors from the LMM and spatial modelling. Refer to Table S 3 of Appendix S 2 for full details. The phylogenetic assignment of the taxa *pmoA*-33, *pmoA*-81 and *pmoA*-130 is indicated in brackets.

	General predictor	Specific predictor	Predictor's effect direction	Residuals spatial dependence (range)
pmoA-33 (USCα)	- Land use	- Grassland, Moorland, Bog	- Negative	Moderate (150 km)
	- Climate	- Mean annual temperature	- Positive	
	- Terrain attribute	- Terrain slope and elevation	- Positive	
pmoA-81 (*Methylocystis/Methylosinus*)	- Land use	- All habitats	- Positive	Moderate (150 km)
pmoA-130 (Cluster 5)	- Land use	- Improved grassland, Woodland	- Negative	Moderate (150 km)

2008; Ramette, 2009; Thies, 2007). The T-RF bins in this study were associated to similar bins (*i.e.* similar bp-length) from several other experiments. The strength of the inferences was further confirmed because T-RFs were generated from soil samples coming from different continents, land-use and soil types but also with samples run on the same fragment analyser, at a different time and by different technicians. See Discussion section for more details. It should be noted that whether cloning/sequencing was performed or not, there will still be a level of uncertainty on the identification of the methanotrophic clades, such as UCSα or *Methylocystis/Methylosinus*. For example, and because of the limitation of the T-RFLP analysis, a *pmoA*-33 T-RF found in a bog (Fig. 2), though not expected, might just as well be a non-USCα individual having a similar restriction site as a USCα member. These taxa were shown to be strongly linked to CH4 consumption across the globe as they are known to be present in aerobic upland soils (Ho et al., 2013; Kolb, 2009; Nazaries et al., 2011; Tate, 2015). Further experimental conditions for T-RFLP are given in the Supplementary Material section of Appendix S 2. As mentioned earlier, a neutral approach was adopted and the label "T-RF" was preferred instead of "taxon".

For clarity, a workflow diagram of the steps involved in modelling methanotroph community structure, relative abundance of the most abundant T-RFs and their environmental drivers is summarised in Fig. 1.

3.2. Drivers of the methanotroph distribution in Scotland

A Spearman's rank correlation analysis identified a range of statistically significant climo-edaphic variables (out of a total of 62) for each of the three selected T-RFs: 46 significant predictors for *pmoA*-33, 57 predictors for *pmoA*-81, and 28 for *pmoA*-130 (Table S 2 in Appendix S 3). These were then included in random forest (RF) analyses to determine the most important environmental drivers of the relative abundance of the T-RFs investigated (Fig. 3). Subsequently, 27 out of 46 predictors were found to significantly influence *pmoA*-33, 24 out of 57 variables for *pmoA*-81 and 14 out of 28 for *pmoA*-130 (Table S 2 of Appendix S 3); thus, accounting for 27–40% of the total variance (Fig. 3). For *pmoA*-33, soil total carbon content, soil moisture and mean annual precipitation were the most important drivers, closely followed by variables such as land use and soil properties like bulk density, total nitrogen, soil minerals and metals (Fig. 3a). Soil water properties (soil aeration, *e.g.* soil drainage and moisture) and land use were the most important drivers of *pmoA*-81 distribution, followed by other spatial (latitude and longitude) and soil properties (Fig. 3b). Finally, the relative abundance of T-RF *pmoA*-130 was mainly driven by soil moisture, soil minerals, altitude and land use (Fig. 3c).

Following RF analysis, the first ten most important and significant variables (Fig. 3) were incorporated into structural equation modelling (SEM) (Fig. 4) to identify the direct, indirect and total effects of the predictors on the distribution of each T-RF. For example, when considering the *pmoA*-33 T-RF (Fig. 4a), soil organic carbon had a total negative effect (Fig. 4a2), that is, it correlated with decreased *pmoA*-33 relative abundance. Meanwhile, organic carbon had a direct negative effect (Fig. 4a1; path coefficient $p = -0.44$) on *pmoA*-33, which was controlled by a direct negative effect ($p = -0.90$) from bulk density (on soil organic carbon). This means that bulk density had an indirect positive effect ($p = (-0.90) \times (-0.44) = 0.40$) (Fig. 4a1). Ultimately, the sum of all direct and indirect effect of bulk density gave a total positive effect of soil bulk density on *pmoA*-33 relative abundance (Fig. 4a2). More theoretical information is found elsewhere (Buttigieg and Ramette, 2014). The same method can be applied to all variables within the SEM diagram. Interestingly, the land uses such as arable land and woodland showed a total positive effect on *pmoA*-33 (Fig. 4a2) while its counterpart *pmoA*-81 exhibited total positive control from bog and moorland (Fig. 4b2). Extra-ordinally, the exact opposite pattern occurred with bog and moorland affecting *pmoA*-33 negatively; and arable land and woodland influencing *pmo*-81 negatively. Similarly, soil organic carbon correlated negatively with *pmoA*-33 (Fig. 4a2) but positively with *pmoA*-81 (Fig. 4b2). A differential effect between *pmoA*-33 and -81 was also observed with soil moisture.

Finally, the SEM network for *pmoA*-130 was a lot simpler although there were not any predictors having a particularly strong total effect on *pmoA*-130 distribution (Fig. 4c2). Bogs had a total positive effect on *pmoA*-130 relative abundance (relative to the other land uses tested), with magnesium being the strongest driver but also other metal ions such as cobalt, nickel and zinc (Fig. 4c2). Noteworthy was the presence of many minerals explaining *pmoA*-130 relative abundance. More specially, these minerals had a direct positive effect ($p = 0.20$) on *pmoA*-130 and an indirect effect *via* metal content ($p = 0.22 \times (-0.76) = 0.17$) on *pmoA*-130 distribution (Fig. 4c1).

3.3. Mapping the distribution of methanotrophs in Scotland

After RF analysis and selection of the most important predictors of methanotrophic T-RF distribution (Figs. 3 and 4), prediction maps were generated using the hybrid spatial modelling approach tested in the present study. Due to the unavailability of continuous surface layers/maps for the variables identified a variety of environmental covariates was used to create the necessary surfaces/maps when upscaling the observation points (see Table 2 and Table S 3 of Appendix S 2). Specifically, these environmental covariates were: land use and land cover maps, climatic variables, digital elevation model and longitude/latitude (see Appendix S 2 for more details). A summary of the environmental factors affecting the relative abundance and distribution of selected *pmoA* T-RFs (as well as their potential phylogenetic assignment) is found in Table 1.

3.3.1. Spatial modelling: estimated fixed and random effects of selected pmoA T-RFs distribution in Scotland

The spatial drivers affected T-RF distribution differentially. For example, land use was identified as a spatial driver for the relative abundance of all three selected T-RFs. When considering the climatic factors, mean annual temperature (MAT) was found to be an important driver for *pmoA*-33 (Table 1 and Table S 3 of Appendix S 2).

The selected T-RFs displayed a moderate spatial dependence of up to ~150 km (Table 1 and Table S 2 in Appendix S 2). Finally, it was important to validate the quality of the spatial models (leave-one-out cross-validation (LOOCV) approach). The mean standardised squared deviation ratio (MSDR) value for the various models was close to 1,

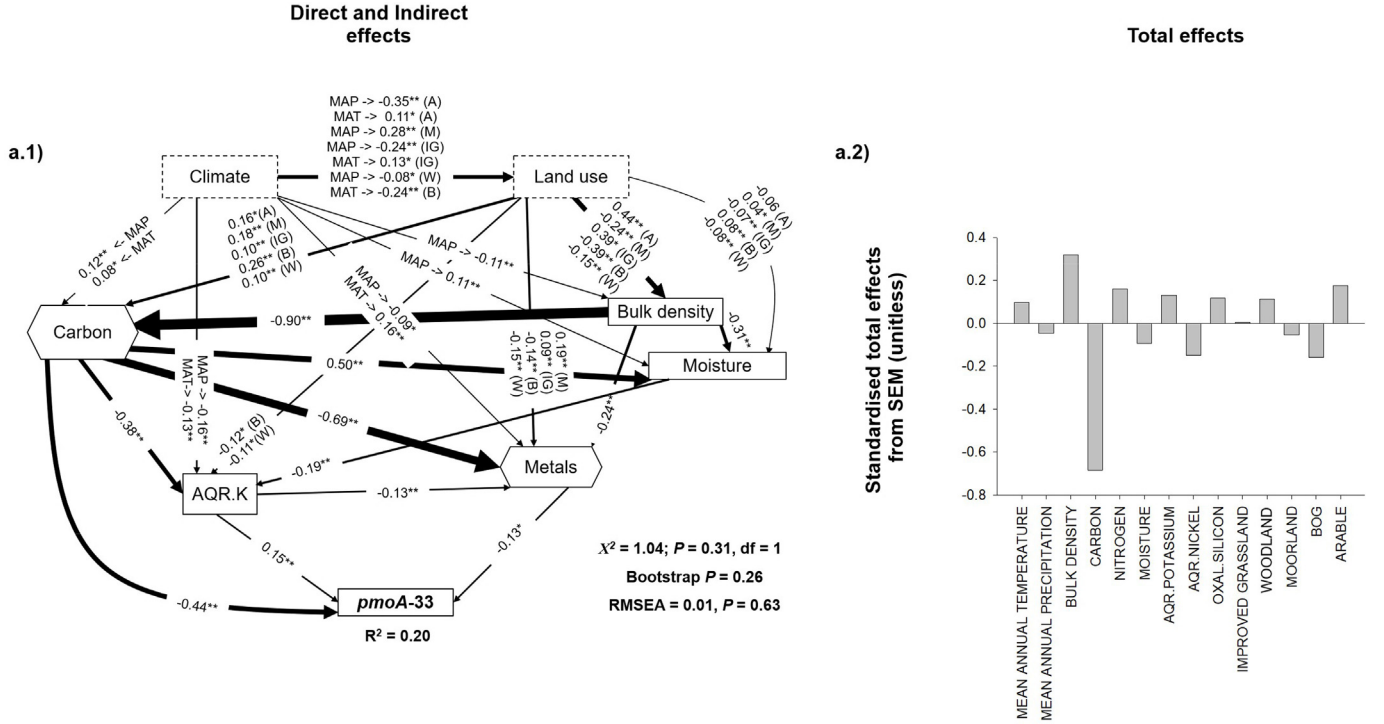

Fig. 4. Direct, indirect and total effects of selected environmental drivers of the distribution of the T-RFs *pmoA*-33 (a), *pmoA*-81 (b) and *pmoA*-130 (c) across Scotland.

Data are derived from the structural equation modelling (SEM) analyses. Panels on the left-hand side (a.1, b.1, c.1) show the output of the path analyses (direct/indirect effects). Panels on the right-hand side (a.2, b.2, c.2) show the total effects, that is, the standardised total effects (direct plus indirect effects). The numbers adjacent to arrows are standardised path coefficients, analogous to partial regression weights, and indicative of the effect size of the relationship. The width of arrows is proportional to the strength of the path coefficients. Hexagons are composite variables while boxes are observable variables. The use of composite variables does not alter the underlying assumptions of SEM, but collapses the effects of multiple conceptually-related variables into a single composite effect, aiding interpretation of the model results (Shipley, 2016). As with other linear models, R^2 indicates the proportion of variance explained and appears alongside every response variable in the model. Model fitness details (χ^2, RMSEA (root mean square error of approximation) and non-parametric Bootstrap parameters) are close to each figure. Only statistically significant paths are displayed. Null hypothesis: the model does not diverge from the dataset. Significance levels are as follows: ns = not significant, *P < 0.05, **P < 0.01 and ***P < 0.001. A list of the predictors included in the SEMs is provided in Appendix S 2. Legend: MAP = mean annual precipitation; MAT = mean annual temperature; A = arable land; B = bog; IG = improved grassland; M = moorland; W = woodland; AQR = Aqua Regia extracted elements; EXCH = exchangeable elements; OXAL = oxalate extractable elements.

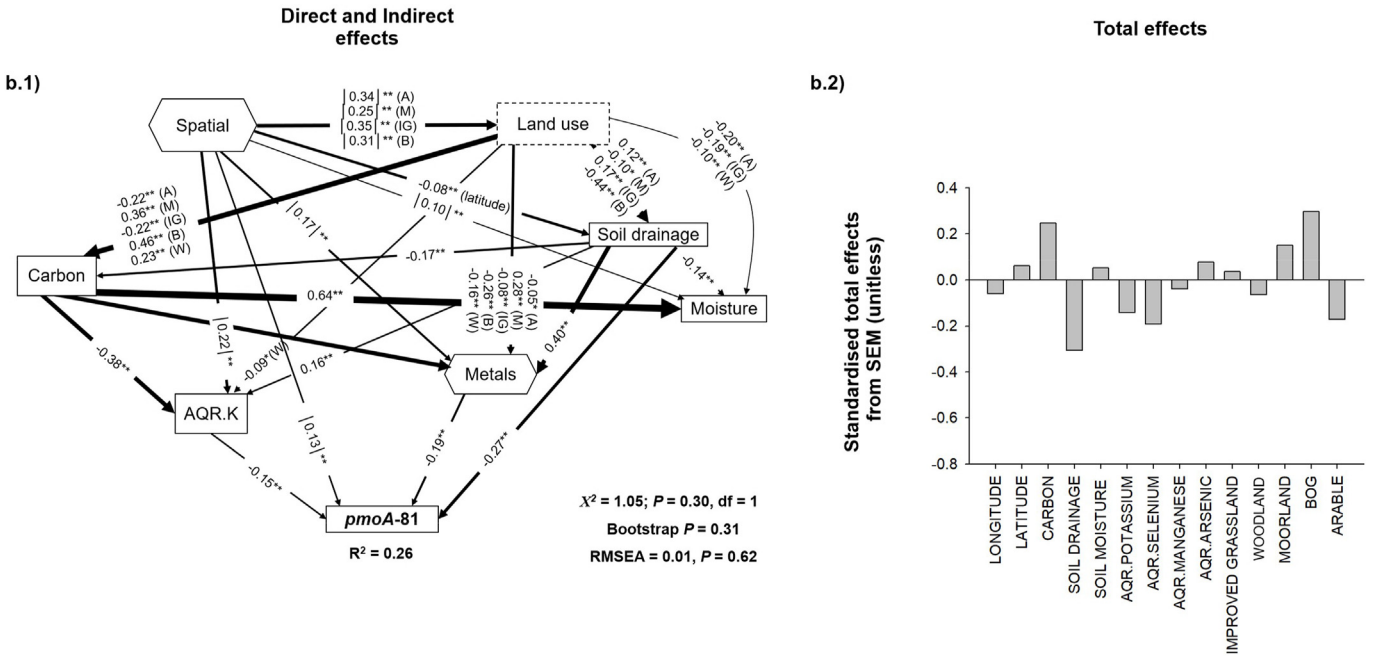

Fig. 4. (*continued*)

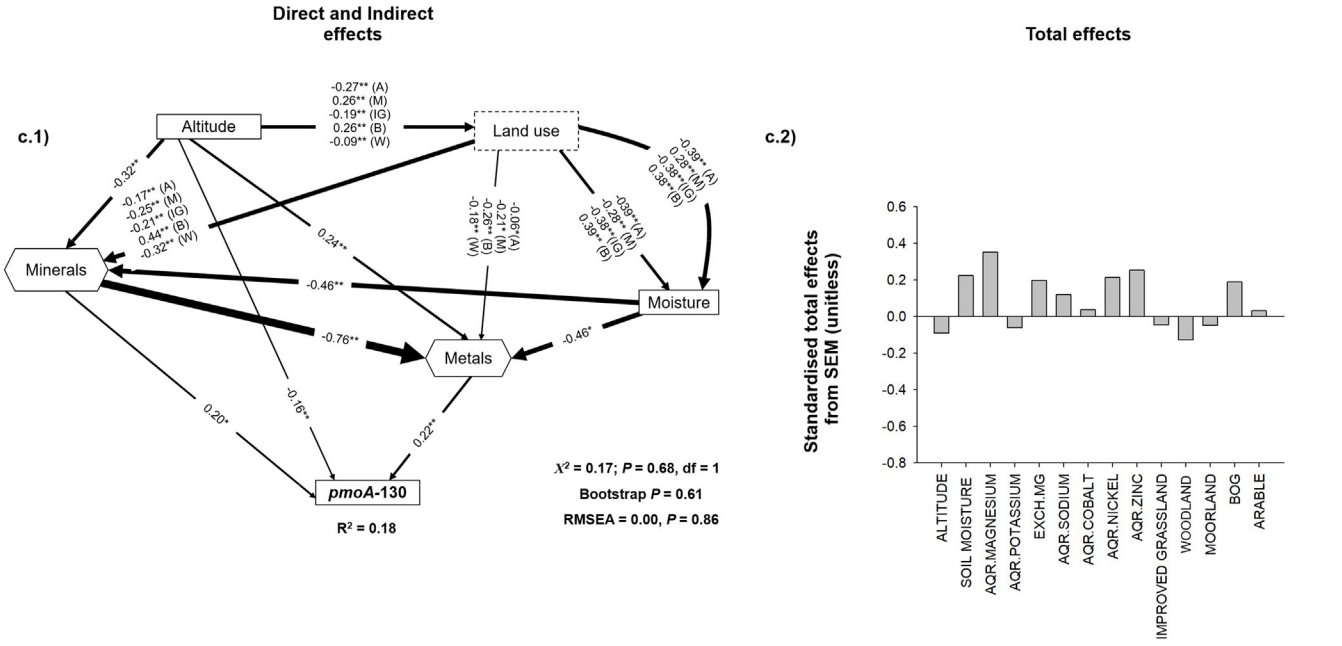

Fig. 4. (*continued*)

Table 2
Estimated random effect terms for the selected T-RFs.

A low nugget-to-sill ratio means that there is a good spatial structure. Cambardella et al. (1994) classified the nugget-to-sill as 'strongly structured spatial dependence' (less than 0.25), 'moderate spatial dependence' (0.25–0.75) and 'weak spatial dependence' (> 0.75).

Model	Nugget (%)	Partial Sill (%)	Sill (%)	Range (m)	Nugget:Sill
pmoA-33	0.0510	0.0338	0.0848	149,787	0.6014
pmoA-81	0.0330	0.0361	0.0691	149,787	0.4776
pmoA-130	0.0058	0.0146	0.0204	149,787	0.2843

Table 3
Summary of the leave-one-out cross-validation (LOOCV) for the *pmoA* relative abundance in Scotland

Mean error (ME) gives an estimate of the bias in the predictions while root mean square error (RMSE) is related to the accuracy of the predictions. A good model has a value close to 0 for ME and RMSE. A mean standardised squared deviation ratio (MSDR) value close to 1 indicates that the prediction variance matches the observed errors. The Lin's concordance correlation coefficient (CCC) indicates how well a plot of the measured *vs.* predicted values follows a 45-degree line. If the CCC value is close to 1 it means that the measured *vs.* predicted values are closely matching each other.

	ME	RMSE	MSDR	CCC
pmoA-33	0.00005	0.2437	1.0005	0.3898
pmoA-81	−0.00017	0.1984	0.9971	0.4748
pmoA-130	−0.00007	0.0864	1.0074	0.6160

which indicated that the models could predict the error properly (Table 3). With respect to Lin's concordance correlation coefficient (CCC), the goodness-of-fit value for the models was greater than 0.38, which indicated a moderate-to-low fit of the data based on the LOOCV approach.

3.3.2. Spatial modelling: methanotroph niche partitioning across Scotland

The hybrid geostatistical modelling approach allowed the spatial prediction of the methanotrophic T-RF relative abundance (Fig. 5), and the corresponding E-BLUP variance maps (equivalent to error variance maps – see Figure S 1 of Appendix S 2). T-RF *pmoA*-81 displayed a higher abundance in the central parts of Scotland and the Highlands (North West), as well as in the North and the Scottish Islands (Fig. 5b). Interestingly, the distribution of T-RF *pmoA*-33 (Fig. 5a) was in some way an opposite and mirror image of that of *pmoA*-81, including a higher abundance in the East Coast of the country, particularly in the Lowlands (South East) (Fig. 5b). In contrast, T-RF *pmoA*-130 reported to have isolated patches of high abundance throughout Scotland (Fig. 5c).

Finally, the T-RF distribution was dependent on the land use (niche partitioning), as suggested by the RF and SEM analyses. Indeed, the relative abundance of the *pmoA*-33 was the highest under arable land and woodland land uses, though *pmoA*-81 was more dominant under wetter soils such as in bog and moorland (Fig. 2). The distribution of these two T-RFs was somehow intermediate under semi-natural and improved grassland, whereas the relative abundance of *pmoA*-130 was constant across the different land uses. The pattern in niche partitioning between *pmoA*-33 and *pmoA*-81 was further confirmed in Fig. 5 by comparing the T-RF distribution maps with the land use map (Fig. 5d). A note of caution though: statistically, the T-RF *pmoA*-33 was negatively correlated with grassland, moorland and bog (Table 1) with P-values of P = 0.024 for grassland, P = 0.001 for moorland and P = 0.000 for bog (see Table S 3 of Appendix S 2). Thus, the opposite effect would be for *pmoA*-33 to positively correlate with woodland habitats which was not statistically supported (Table S 3 of Appendix S 2). It is commonly established that methanotrophs of the UCSα clade are consistently found across the world in a range of habitats/upland soils, from China, Thailand, Japan, Brazil, Canada, New Zealand, USA, Germany (Degelmann et al., 2010; Han et al., 2009; Lima et al., 2014; Martineau et al., 2014; Narihiro et al., 2011; Nazaries et al., 2011; Wu et al., 2013). Further, members of the UCSα clade were found in many woodlands across the globe and are usually found to correlate with atmospheric CH_4 sink (Knief, 2015). However, this was not explicitly supported from the LMM analysis but one could assume that this is a highly plausible scenario within the present dataset.

While the three selected T-RFs were all assumed to be representatives of α-proteobacterial methanotrophs, another relatively abundant T-RF (*pmoA*-247) was found in ~3% of all land uses (Fig. 2 and Table S 1 of Appendix S 2). As with the other dominant T-RFs, this *pmoA*-247 may also be assumed to be identified as a representative of γ-proteobacterial methanotrophs, more specifically *Methylococcus* spp. (Nazaries et al., 2011; Singh et al., 2007). The T-RF *pmoA*-247 also showed to significantly influence the methanotrophic community

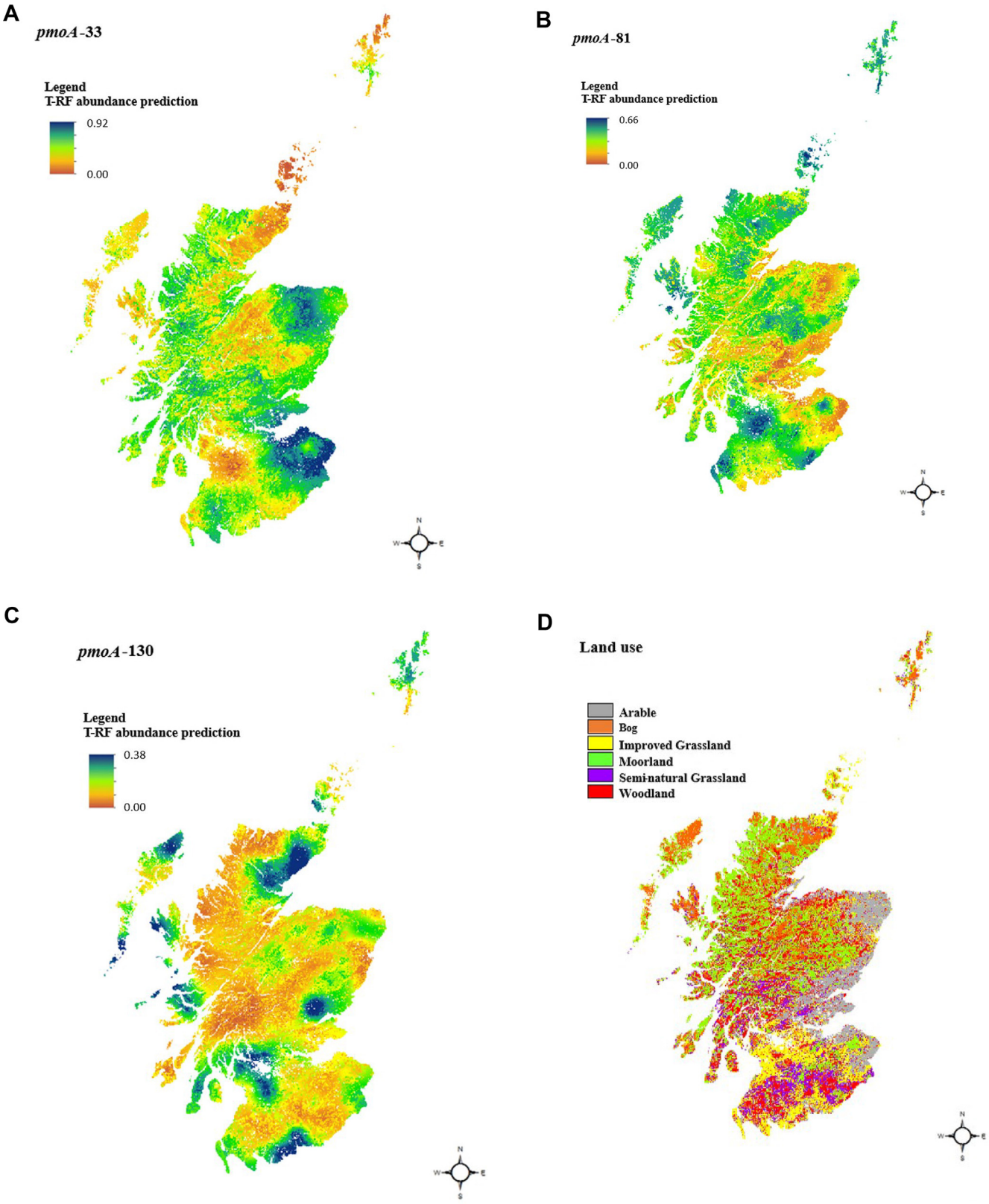

Fig. 5. Spatial distribution of selected *pmoA* T-RF relative abundance across Scotland.
The T-RFs presented are: *pmoA*-33 (a), *pmoA*-81 (b) and *pmoA*-130 (c). A map of the six broad land use types is also included in the panel (d). An estimate of uncertainty of the predicted maps was given in the form of prediction variance for each grid as depicted in Figure S 3 of Appendix S 3.

Table 4
Influence of individual T-RFs loadings on the methanotrophic community structure.

Data show the T-RF loadings (a.k.a. eigen-values/factors, coefficients, weights) for the first ten (a) and last ten (b) T-RFs of the first three PC axes following T-RFLP analyses. Loadings were sorted as "smallest-to-largest" for each PC axe, and for both top/last 10 T-RFs. The most influential methanotrophic T-RF can be differentiated by a colour. The sign of a loading does not affect its "direction" (*i.e.* positive/negative coefficient), meaning, only the loading absolute value is indicative of a T-RF effect (Buttigieg and Ramette, 2014). Loadings highlighted in yellow and bold type represent T-RFs with a significant impact (loading ≥ |0.30|) on the community structure. Red font shows T-RFs with a lesser influence (loading ≈ |0.15| to |0.30| on community structure while black font was for non-significant loadings (loading ≤ |0.15|).

	PC1-sorted loadings				PC2-sorted loadings				PC3-sorted loadings			
		PC1	PC2	PC3		PC1	PC2	PC3		PC1	PC2	PC3
(a) Top 10 T-RFs	*pmoA*-33	-0.358	-0.213	-0.111	*pmoA*-359	-0.097	-0.317	0.248	*pmoA*-81	0.339	0.185	-0.206
	pmoA-48	-0.151	-0.104	0.119	*pmoA*-247	-0.004	-0.315	0.248	*pmoA*-84	0.096	0.100	-0.160
	pmoA-128	-0.149	-0.209	0.196	*pmoA*-33	-0.358	-0.213	-0.111	*pmoA*-266	-0.004	0.159	-0.128
	pmoA-36	-0.145	-0.009	-0.030	*pmoA*-128	-0.149	-0.209	0.196	*pmoA*-77	0.082	0.118	-0.118
	pmoA-63	-0.131	0.244	0.223	*pmoA*-481	0.171	-0.133	0.261	*pmoA*-54	-0.013	0.086	-0.116
	pmoA-106	-0.128	0.216	0.178	*pmoA*-48	-0.151	-0.104	0.119	*pmoA*-33	-0.358	-0.213	-0.111
	pmo-51	-0.105	0.185	0.079	*pmoA*-149	0.028	-0.103	0.219	*pmoA*-31	-0.072	0.008	-0.094
	pmoA-103	-0.101	0.208	0.088	*pmoA*-434	0.179	-0.083	0.182	*pmoA*-116	0.044	0.061	-0.061
	pmoA-359	-0.097	-0.317	0.248	*pmoA*-123	-0.033	-0.067	0.114	*pmoA*-36	-0.145	-0.009	-0.030
	pmo-99	-0.096	0.129	0.176	*pmoA*-130	0.283	-0.061	-0.016	*pmoA*-294	0.052	0.059	-0.027
(b) Last 10 T-RFs	*pmoA*-84	0.096	0.100	-0.160	*pmoA*-266	-0.004	0.159	-0.128	*pmoA*-434	0.179	-0.083	0.182
	pmoA-311	0.098	0.010	0.062	*pmoA*-71	-0.026	0.173	0.048	*pmoA*-128	-0.149	-0.209	0.196
	pmoA-481	0.171	-0.133	0.261	*pmoA*-51	-0.105	0.185	0.079	*pmoA*-216	-0.060	0.234	0.207
	pmoA-434	0.179	-0.083	0.182	*pmoA*-81	0.339	0.185	-0.206	*pmoA*-144	0.090	0.037	0.214
	pmoA-433	0.258	-0.007	0.084	*pmoA*-109	-0.056	0.205	0.056	*pmoA*-149	0.028	-0.103	0.219
	pmoA-432	0.260	-0.012	0.101	*pmoA*-103	-0.101	0.208	0.088	*pmoA*-63	-0.131	0.244	0.223
	pmoA-130	0.283	-0.061	-0.016	*pmoA*-106	-0.128	0.216	0.178	*pmoA*-470	0.367	-0.028	0.245
	pmoA-88	0.334	0.052	0.024	*pmoA*-216	-0.060	0.234	0.207	*pmoA*-359	-0.097	-0.317	0.248
	pmoA-81	0.339	0.185	-0.206	*pmoA*-63	-0.131	0.244	0.223	*pmoA*-247	-0.004	-0.315	0.248
	pmoA-470	0.367	-0.028	0.245	*pmoA*-190	-0.052	0.275	0.082	*pmoA*-481	0.171	-0.133	0.261

assemblage (see principal component analysis in Supporting Results and Figure S 1 of Appendix S 3). Yet, *pmoA*-247 relative abundance remained uniform across the Scottish landscape (Fig. 2 and Table S 1 of Appendix S 2).

3.4. Mapping the methanotrophic community assemblage across Scotland

The spatial modelling was also conducted at the whole community level by producing maps of the methanotrophic community assemblage based on the entire *pmoA* T-RFLP dataset, that is, based on the application of principal component analysis (PCA) as a proxy for community composition of Appendix S 3).

The indicators of modelling quality were similar to those observed for the individual T-RF relative abundance (Table S 1 to Table S 3 of Appendix S 3). Concisely, the LMM fixed effects identified land use as the main predictor of PC1, while PC2 was more related to land use, climatic properties (mean annual temperature) and spatial attributes (terrain slope and elevation). The random effects were identified as moderate-to-weak with a spatial dependence of up to ~150 km. See Appendix S 3 for more detailed description.

As an added value, the PCA loadings (also called weights or eigenvalues) for each (three) PC axis were inspected individually (Table 4). While the cumulative T-RF abundance was 63%, the category "Other" accounted for ≈24% of *pmoA* distribution and through-out the different land uses. This category consisted of a long list of rare T-RFs all < 1% of the of the total relative abundance (see Table S 1 in Appendix S 2). Finally, this left ≈13% of mid-range T-RFs not included into the statistical analyses. In parallel, mapping of the community structure was only applied to the first three PC axes. This is because the scree plot in Figure S 1d showed a "broken elbow" at the third axis (38% of total inertia, cumulatively), meaning that most of the variance was accounted for. The analysis could have been extended to include PC4 and PC5 as they still contained ≈20% of the variation in methanotroph assemblage.

The PCA-produced loadings of the top and last 10 T-RFs, that is, the most and least (respectively) influential T-RFs on the community structure were identified (Table 4). On the first PC, *pmoA*-33 and *pmoA*-81 had the highest loadings [*i.e.* loading ≥ |0.30| – see Buttigieg and Ramette (2014)] of the top/lowest 10 T-RFs (Table 4). When considering the raw data (*i.e.* the signed loadings – see Table 4), the present results agreed with those in section 3.2 and thus further highlights the existence of a niche partitioning between these two T-RFs. Other T-RFs (*pmoA*-88 and *pmoA*-470 – see last 10 T-RFs in (Table 4b) were also found to significantly alter the methanotrophic community, although their relative abundance was last then 1% (see Fig. 2). On the second PC, *pmoA*-359 and *pmoA*-247 (methanotrophs of yet unknown identity) had the highest loadings (Table 4a) although most of the other T-RFs still showed intermediate (loading ≈ |0.15| to |0.30|) effects on the methanotrophic community structure. Finally, on the third PC axis, many of the T-RFs showed again an intermediate effect on community

structure (Table 4).

4. Discussion

This study provides robust empirical evidence on the environmental drivers of methanotrophic community and dominant taxa and, as such, suggests that the distribution of methanotrophs across an extent of land can be predicted at a large geographical scale and is mainly driven by land use and soil properties (*e.g.* soil moisture or organic matter content). This information can then be used for mapping T-RF distribution (*e.g.* methanotrophs) at different geographical scales (*e.g.* regional, national) with possible implications for the understanding of ecosystem functioning, such as global CH_4 emission budgets with the aim to implement appropriate policies and management practices in relation to current global change challenges. Moreover, understanding the impact that environment has on the relative abundance of keystone taxa can help better identify the ecological mechanisms of competition and their adaptation to specific niches. To our knowledge, this is the first attempt to map at a large geographical scale a specialised microbial group using both T-RF distribution and community assemblage, as well as explicitly including spatial and terrain attributes as part of the geostatistical interpolation. Given the strong link between methanotrophic communities and net CH_4 emissions (Ho et al., 2013; Kolb, 2009; Nazaries et al., 2011; Tate, 2015), this approach may help unravel the potential consequences of land management and global environmental changes on biogeochemical cycles (*e.g.* CH_4 emissions), as well as by improving Earth's ecosystem models.

4.1. Environmental drivers of the dominant methanotrophs: insight into microbial niche partitioning

The RF and SEM analyses identified land use (woodland, bog and arable land), soil properties (organic matter), minerals, metal ions, climate (rainfall) and spatial (longitude and latitude) attributes as main predictors of the relative abundance of T-RFs *pmoA*-33 and *pmoA*-81, while *pmoA*-130 relative abundance was only predicted by soil properties (minerals and metals) and altitude (Figs. 3 and 4). Since no cloning/sequencing was performed, these clades of methanotrophs potentially belong to taxa related to, respectively, USCα, *Methylocystis/Methylosinus*, and methanotroph's Cluster 5 (Table 1 and Nazaries, 2011; Nazaries et al., 2013b). The high incidence of the "alleged" USCα T-RF (*pmoA*-33 – see Results section) in this manuscript might seem unexpected in bogs (site of high methanogenesis activity) yet, a recent study on the presence of methanotrophs in bog revealed a high proportion of USCα and *Methylocystis* spp. methanotroph (Danilova et al., 2016) but also in other acidic wetlands such as *Sphagnum* mosses, swamps, fens *etc*. where atmospheric CH_4 oxidation occurred (Dedysh, 2009; Kolb and Horn, 2012; Whalen, 2005). The reverse argument is also valid with the elevated incidence of *Methylocystis* spp. in bogs.

These taxa are ubiquitous worldwide and reported in a wide range of habitats/upland soils across the globe such as in China, Thailand, Japan, Brazil, Canada, New Zealand, USA, Germany (Wu et al., 2013; Narihiro et al., 2011; Degelmann et al., 2010; Lima et al., 2014; Han et al., 2009; Martineau et al., 2014 and referencies in Nazaries et al., 2011). Unexpectedly, SEM revealed the relative abundance of the three most dominant T-RFs to be mostly explained by land use and a narrow range of environmental (climo-edaphic) properties as discussed below:

(1) Throughout the Results section, the T-RFs *pmoA*-33 and *pmoA*-81 had similar drivers, but with opposite effects (see Fig. 4a2; 4b2; 4c2). For example, T-RF *pmoA*-33 was directly influenced by soil organic matter, moisture and land use (positively and negatively related to arable land and bog, respectively) whereas the same drivers were found to alter the relative abundance of T-RF *pmoA*-81 in an opposite manner suggesting niche partitioning. This disparity concurs with previous findings of the impact of land use and moisture on methanotrophs (Tate, 2015).

(2) In a recent publication, Knief (2015) proposed that habitat specificity of cultivated and uncultivated methanotrophs could be classified as "generalists" and "specialists" depending on the taxonomic resolution considered (genus *vs.* species). According to the author, habitat specificity with members of the USCα clade was predominantly "soil with weak CH_4 supply" (*i.e.* specialist methanotrophs), whereas *Methylocystis/Methylosinus* were found in a wide range of habitats (*i.e.* generalist methanotrophs). Presently, while T-RF *pmoA*-33 might be described as specialist methanotroph because of its high affinity toward atmospheric CH_4, this was not the case here because USCα co-existed (see next section) with *Methylocystis/Methylosinus* in most land uses, including bogs and other wetlands (see Fig. 2). This observation is in line with the findings described in the previous section.

(3) The USCα *vs. Methylocystis/Methylosinus* competition for habitat dominance (or niche partitioning) between these two T-RFs could be compared to a co-existence or equilibrium. Recently, Esson et al. (2016) discovered that both γ- and α-proteobacterial methanotrophs could co-dominate atmospheric CH_4 oxidation in soils under peat bog. In addition, this pattern of co-existence was also observed under grassland site within members of the γ-proteobacterial methanotrophs with an apparent lack of α-proteobacterial methanotrophs (Judd et al., 2016). Here, this pattern was "reversed", *i.e.* soils contained two co-dominant α-proteobacterial methanotrophs (USCα *vs. Methylocystis/Methylosinus*). It is thus possible that the two groups of methanotrophs (USCα *vs. Methylocystis/Methylosinus*) slowly adjust their relative abundance according to changes in environmental conditions such as climate (MAP and MAT) for USCα (Fig. 4a), geography (latitude/longitude) for *Methylocystis/Methylosinus* members (Fig. 4b) and altitude for *pmoA*-130 (Fig. 4c). Ergo, *pmoA*-33 and *pmoA*-81 could be both defined as generalists, while *pmoA*-130 could behave as a specialist. This "specificity" of *pmoA*-130 could be explained due to the relatively better quality of its descriptors (lowest nugget:sill ratio; lowest RMSE; highest CCC – see Tables 2 and 3) when producing the *pmoA*-130 map. The lack of precision of the *pmoA*-33 and *pmoA*-81 maps may also be explained from comparison against the mirrored descriptors of *pmoA*-130 (also, see Fig. 5a and b). Maybe it would help close the knowledge gap for better understanding the co-existence of methanotrophs with other less abundant T-RFs. Another analysis could focus on the "less important" PC axes as some unexplained variance was left in the last two PCs (PC-4 and PC-5), as indicated in Figure S 1c in Appendix S 3. The rarest T-RFs might be more represented within the last two PCs (or even PC3). More work is needed to better understand the concept of niche partitioning such as "why are there few common species but many unique ones" as investigated in other branches of ecology (Kurm et al., 2017; Siqueira et al., 2012). The outcome of the present study further suggests that T-RF-specific niches could be identified at the regional scale and that it can be used to predict the distribution of methanotrophs at a large geographical scale.

(4) The differential effect of soil moisture and land use on the distribution of methanotrophic T-RFs is further supported by a recent report form Ho et al. (2013) which proposed that low- and high-affinity methanotrophs compete for CH_4 and O_2 within the soil profile, where high-affinity methanotrophs such as USCα dominate under low CH_4 concentration but high O_2 levels, whereas low-affinity methanotrophs such as *Methylocystis/Methylosinus* thrive under high CH_4 and low O_2. Concurrently, this study suggests that under conditions where soil moisture is low (*e.g.* under arable and woodland soils) and competition for substrate is high, then USCα would dominate, while when soil moisture becomes high (*e.g.* under highly organic and wet bog and moorland soils) then *Methylocystis/Methylosinus* methanotrophs would thrive. This is possibly another example that soil moisture might be the strongest predictor of

methanotrophic biogeography. For instance, Judd et al. (2016) observed in grassland sites that a precipitation gradient was correlated with co-dominance of methanotrophic clades. The RF modelling suggested that the top four environmental variables explaining the methanotrophic T-RF distribution across Scotland were soil organic matter, minerals, moisture and land use. The SEM also confirmed that soil moisture was positively related to land uses such as moorland and bog (Fig. 4), where the moisture content is generally high. The opposite was true for other land uses with lower soil moisture (*i.e.* grassland, arable land and woodland). SEM further supported that soil moisture had a strong, direct and total effect on the methanotrophic T-RFs. Interestingly, soil metal ion content was found to have a direct effect on the three most abundant T-RFs (Fig. 4a1, b1, c1). This has not been identified in field experiments before. Many biological enzymes need metal ions in their active site (Pratscher et al., 2018). This is particularly true for monooxygenases like the two forms (particulate and soluble) of methane monooxygenases (MMO). Both enzymes contain a metal (iron and/or copper) ion core (Lieberman and Rosenzweig, 2004; Semrau et al., 2010). It is also now well-known that pMMO enzymatic activity is regulated by soil copper concentration (a.k.a "copper-switch") (Nielsen et al., 1997; Semrau et al., 2010; Trotsenko and Murrell, 2008) as well as by the secretion in the environment of a copper-binding protein (a.k.a. chalkophore, or copper-shuttle) called methanobactin (Dassama et al., 2017; Kim et al., 2005; Semrau et al., 2010). In the present study, the direct role of copper was not established as a statistically significant driver of T-RF relative abundance in Scotland. Pratscher and colleges (2018) identified a range of features derived from the draft genome of USCα, and in particular, the presence of membrane-bound transporters of metal ions (zinc, cadmium, copper, magnesium). This is matching with the SEM analyses which highlighted that several metal ions were considered important predictors of *pmoA* T-RFs, especially *pmoA*-130 (Cluster 5 methanotrophs). Still, recent observations emphasised the potential involvement of metal ions ("metal-switch") in methanotrophy (Dassama et al., 2017; Semrau et al., 2018). This provides novel evidence that metal concentration in soil can be a key driver of methanotrophic T-RF distribution at a large geographical scale.

4.2. A hybrid geostatistical approach to spatial modelling: mapping and predicting the biogeography of methanotrophs at the Scottish national scale

The current work provides the first attempt to produce spatial prediction maps of keystone specialised microbes (methanotrophs) while accounting for methanotrophic community assemblage at a regional scale using a hybrid geostatistical modelling approach. Contrary to other approaches, the spatial interpolation presented here combines statistical analysis (Random Forest) of climo-edaphic predictors with spatial predictions based on linear-mixed modelling as it allows for the inclusion of explicit spatial and terrain attributes, as well as the production of error (E-BLUP) variance maps (Figure S 1 of Appendix S 2). Using LMM permits the incorporation of environmental drivers of pmoA genes as covariates (deterministic component of the model) while unexplained variability (residuals) are modelled as spatially correlated random effect terms. Therefore, the use of LMM is more advantageous for modelling T-RF distribution compared to traditional geostatistical approaches (Lark et al., 2006), especially in modelling highly heterogeneous environments such as across Scotland.

A similar approach to study methanotroph distribution/composition was attempted before (Krause et al., 2013) but this was at a much lower scale (10 × 10 m grid) and with a different geostatistical approach. Most studies related to the prediction of methanotroph distribution using *pmoA* genes, with respect to spatial variability modelling, were carried out for methanotrophic diversity indexes such as richness, evenness and Shannon index (*e.g.* Krause et al., 2013). A limitation to this hybrid modelling approach was its relatively low-resolution power. Looking at SEM results, the R^2 values were, respectively, 0.20 for *pmoA*-33, 0.26 for *pmoA*-81 and 0.18 for *pmoA*-130 (Table 3). This suggested a low level of performance to capture the variables effects on T-RFs, while the findings of Krause et al. (2013) estimated better resolution powers. It is important to understand that the LMMs used here were developed as a linear relationship between environmental covariates. One possible reason for less power of our fitted LMM may be due to non-linear relationship between methanotrophs and environmental covariates. Therefore, it is encouraged that future studies may adapt the same framework but also try different model fitting processes such as data mining and machine learning approaches. These approaches for mapping soil properties are widely used in the scientific literature as in Hengl et al. (2017).

The present analysis is similar to, yet, different from a study by Bru et al. (2011). The main points of difference (or similarity) are summarised in Supplementary Table S 4. The ecosystem functions investigated were involved in the nitrogen- and carbon-cycle (respectfully, N_2O *vs.* CH_4 greenhouse fluxes). While both studies shared similar area range and grid size, the results from the variogram analyses suggested a moderate spatial dependence (150 km) in methanotrophic community composition and distribution whereas the auto-correlations of the N-cycle response variables were between 22 and 739 km depending on the gene modelled. Spatial heterogeneity of microbes under varying ecosystems is also dependent on vegetation gradient, sampling scale and taxonomic resolution (Ettema and Wardle, 2002). In the present study, and as in Bru et al. (2011), there was a strong correlation between habitats and ecosystem functional groups, owing to the congruent observations of methanotroph niche-partitioning (Figs. 3–5 and Table 4 of main text; and Figure S 2 of Appendix S 3). This implies that, as a whole, microbial biogeography has a positive taxa-area relationship (Horner-Devine et al., 2004), although selection and isolation by geographical distance (*e.g.* endemism) is also relevant in many cases (Ramette and Tiedje, 2007). Nonetheless, the present study suggests that the spatial distribution of methanotrophs follows biogeographical patterns similar to those observed for broad microbial phyla, such as Bacteria or Fungi (Constancias et al., 2015; Dequiedt et al., 2011; Fierer and Jackson, 2006; Tedersoo et al., 2014).

There were other points of difference between Bru et al. (2011) and our approach, particularly related to sample collection, data generation and analysis but more importantly, data preparation for geostatistical interpolation (Supplementary Table S 4). Although it is difficult to assess which approach was the best (based on data availability), it was obvious that each method had pros and cons. An interesting methodology would be to run each response variable with both statistical frameworks and "quantify" the better approach.

4.3. Broader implications of microbial modelling for predicting terrestrial ecosystem functioning

The methanotrophs detected in the present study have been recognised as major contributors of atmospheric (trace amount) CH_4 oxidation due to their high affinity to CH_4 (Henckel et al., 2000; Knief, 2015). Other studies established links between the CH_4-oxidising potential of a land and its associated compositional assemblage of soil methanotrophic communities (Nazaries et al., 2013b, 2011; Singh et al., 2007). Particularly, an opposite distribution of USCα, *Methylococcus*-related and/or *Methylocystis/Methylosinus* across various land uses was previously reported by Nazaries et al. (2013b, 2011) and Knief (2015). The focus of the actual study was to examine the biogeography of methanotrophs and hence CH_4 fluxes were not measured. However, our data supports "niche partitioning" amongst dominant taxa since the distribution of the *pmoA* T-RF was consistent with the expected net CH_4 flux under those land uses, *e.g.* high proportion of USCα methanotrophs under woodland (where CH_4 oxidation is generally high) or low proportion under bog (where CH_4 emissions are usually high) (see Table 1,

Fig. 2 and Table S 2 of Appendix S 3).

There is a global trend in using process-based models to evaluate changes in greenhouse gas emissions under different agronomic practices or climatic scenarios (Butterbach-Bahl et al., 2001; Li et al., 1996; Saggar et al., 2007). Yet, most process-based models do not have any soil microbial module even though those soil microbes are fundamental to these processes (Nazaries et al., 2013a). Therefore, the predictive maps produced in the current study may lay foundation for future work to combine soil microbial data (taxonomy and traits) with process-based models for future application in a spatio-temporal context, such as trait-based biogeography (Green et al., 2008; Powell et al., 2015b).

5. Conclusion and perspectives

The present study provides a large geographical scale population and community distribution pattern of methane-oxidising bacteria. Adopting an emerging modelling approach, prediction maps (and their corresponding error variance) of the relative abundance of dominant methanotrophic T-RFs were produced following the identification of main environmental drivers of methanotrophic distribution across Scotland. Interestingly, the distribution of methanotrophs linked to the clades USCα, *Methylocystis/Methylosinus* and Cluster 5 was predicted by different environmental variables suggesting a clear niche partitioning. Importantly, soil properties (carbon, moisture, minerals and metals) and land use were the main drivers of niche partitioning and competition between the dominant groups of methanotrophs. Two different modelling approaches (RF/SEM analyses and LMM) were applied and both techniques gave very similar results, although SEM was more powerful as it allowed to construct paths to estimate the direct, indirect and total effects of the various predictors on the response variables. The hybrid geostatistical approach tested in the present study provides a novel framework to identify and model the distribution of microbial communities along spatial and environmental gradients, which is presently a key constraint in incorporating microbial data into simulation models and land management policies. While the study did not allow to test the direct relationship between methanotrophic community and net methane fluxes, this work paves the first step for a better understanding of atmospheric methane mitigation and environmental properties regulating this process.

6. Data accessibility statement

Readers are welcome to access the complete set of climo-edaphic properties and taxonomic data available for the present study. This can be found in the Supplementary Material section, particularly in Appendix S 1.

Conflicts of interest

None.

Appendix S 1: NSIS_2 dataset used in the present study. Details on the sampling design and analytical procedures are provided in Lilly et al. (2010a, 2010b).

Appendix S2: Details of the molecular fingerprinting analysis and hybrid geostatistical modelling approach for mapping selected methanotrophic T-RFs (*pmoA*-33, *pmoA*-81 and *pmoA*-130).

Appendix S 3: Details of the hybrid geostatistical modelling approach for mapping selected methanotrophic community assemblage (PC1, PC2 and PC3).

Acknowledgements

The authors would like to thank the NSIS team at the James Hutton Institute (Aberdeen, Scotland) for their effort in soil sampling and processing. We also acknowledge the Analytical Department at JHI for performing the soil physico-chemical analyses and Allan Lily for providing map details. We are most grateful for Nadine Thomas' hard work on the molecular analyses. Finally, we want to thank the two anonymous reviewers and the editor for their input in this study. M.D-B. acknowledges support from the Marie Sklodowska-Curie Actions of the Horizon 2020 Framework Programme H2020-MSCA-IF-2016 under REA grant agreement n° 702057. The B.K.S. team was supported by Australian Research Council grants (DP 170104634).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.soilbio.2018.08.014.

References

Bourne, D.G., McDonald, I.R., Murrell, J.C., 2001. Comparison of pmoA PCR primer sets as tools for investigating methanotroph diversity in three Danish soils. Applied and Environmental Microbiology 67, 3802–3809.

Bru, D., Ramette, A., Saby, N.P.A., Dequiedt, S., Ranjard, L., Jolivet, C., Arrouays, D., Philippot, L., 2011. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. The ISME Journal 5, 532–542. https://doi.org/10.1038/ismej.2010.130.

Butterbach-Bahl, K., Stange, F., Papen, H., Li, C., 2001. Regional inventory of nitric oxide and nitrous oxide emissions for forest soils of Southeast Germany using the biogeochemical model PnET-N-DNDC. Journal of Geophysical Research 106, 34155–34165.

Buttigieg, P.L., Ramette, A., 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiology Ecology 90, 543–550. https://doi.org/10.1111/1574-6941.12437.

Cambardella, C.A., Moorman, T.B., Parkin, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., Konopka, A.E., 1994. Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America 58, 1501–1511.

Conrad, R., 2009. The global methane cycle: recent advances in understanding the microbial processes involved. Environmental Microbiology Reports 1, 285–292. https://doi.org/10.1111/j.1758-2229.2009.00038.x.

Conrad, R., 2007. Microbial ecology of methanogens and methanotrophs. In: Advances in Agronomy. Academic Press, pp. 1–63. https://doi.org/10.1016/S0065-2113(07)96005-8.

Constancias, F., Saby, N.P.A., Terrat, S., Dequiedt, S., Horrigue, W., Nowak, V., Guillemin, J.P., Biju-Duval, L., Chemidlin Prévost-Bouré, N., Ranjard, L., 2015. Contrasting spatial patterns and ecological attributes of soil bacterial and archaeal taxa across a landscape. Microbiology (Washington D C) 518–531. https://doi.org/10.1002/mbo3.256.

Culman, S.W., Bukowski, R., Gauch, H.G., Cadillo-Quiroz, H., Buckley, D.H., 2009. T-REX: software for the processing and analysis of T-RFLP data. BMC Bioinformatics 10.

Culman, S.W., Gauch, H.G., Blackwood, C.B., Thies, J.E., 2008. Analysis of T-RFLP data using analysis of variance and ordination methods: a comparative study. Journal of Microbiological Methods 75, 55–63. https://doi.org/10.1016/j.mimet.2008.04.011.

Danilova, O.V., Belova, S.E., Gagarinova, I.V., Dedysh, S.N., 2016. Microbial community composition and methanotroph diversity of a subarctic wetland in Russia. Microbiology 85, 583–591.

Dassama, L.M.K., Kenney, G.E., Rosenzweig, A.C., 2017. Methanobactins: from genome to function. Metallomics 9, 7–20. https://doi.org/10.1039/C6MT00208K.

Dedysh, S.N., 2009. Exploring methanotroph diversity in acidic northern wetlands: molecular and cultivation-based studies. Microbiology 78, 655–669. https://doi.org/10.1134/S0026261709060010.

Degelmann, D.M., Borken, W., Drake, H.L., Kolb, S., 2010. Different atmospheric methane-oxidizing communities in european beech and Norway spruce soils. Applied and Environmental Microbiology 76, 3228–3235. https://doi.org/10.1128/AEM.02730-09.

Delgado-Baquerizo, M., Gallardo, A., Covelo, F., Prado-Comesaña, A., Ochoa, V., Maestre, F.T., 2015. Differences in thallus chemistry are related to species-specific effects of biocrust-forming lichens on soil nutrients and microbial communities. Functional Ecology 29, 1087–1098. https://doi.org/10.1111/1365-2435.12403.

Delgado-Baquerizo, M., Giaramida, L., Reich, P.B., Khachane, A.N., Hamonts, K., Edwards, C., Lawton, L.A., Singh, B.K., Brophy, C., 2016a. Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. Journal of Ecology 104, 936–946. https://doi.org/10.1111/1365-2745.12585.

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar, D., Berdugo, M., Campbell, C.D., Singh, B.K., 2016b. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nature Communications 7, 1–8. https://doi.org/10.1038/ncomms10541.

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Trivedi, P., Osanai, Y., Liu, Y.-R., Hamonts, K., Jeffries, T.C., Singh, B.K., 2016c. Carbon content and climate variability drive global soil bacterial diversity patterns. Ecological Monographs 86, 373–390. https://doi.org/10.1002/ecm.1216.

Dequiedt, S., Saby, N.P.A., Lelievre, M., Jolivet, C., Thioulouse, J., Toutain, B., Arrouays, D., Bispo, A., Lemanceau, P., Ranjard, L., 2011. Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. Global Ecology and Biogeography 20, 641–652. https://doi.org/10.1111/j.1466-8238.2010.00628.x.

Dumont, M.G., Lüke, C., Deng, Y., Frenzel, P., 2014. Classification of pmoA amplicon

pyrosequences using BLAST and the lowest common ancestor method in MEGAN. Frontiers in Microbiology 5, 1–11. https://doi.org/10.3389/fmicb.2014.00034.
Esson, K.C., Lin, X., Kumaresan, D., Chanton, J.P., Murrell, J.C., Kostka, J.E., 2016. Alpha- and gammaproteobacterial methanotrophs codominate the active methane-oxidizing communities in an acidic boreal peat bog. Applied and Environmental Microbiology 82, 2363–2371. https://doi.org/10.1128/AEM.03640-15.
Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. Trends in Ecology & Evolution 17, 177–183. https://doi.org/10.1016/S0169-5347(02)02496-5.
Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., De Beer, D., 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature 464, 543.
Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 103, 626–631. https://doi.org/10.1073/pnas.0507535103.
Grace, J.B., 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press.
Green, J.L., Bohannan, B.J.M., Whitaker, R.J., 2008. Microbial biogeography: from taxonomy to traits. Science 320, 1039–1043.
Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils. Environmental Microbiology 13, 1642–1654. https://doi.org/10.1111/j.1462-2920.2011.02480.x.
Han, B., Chen, Y., Abell, G., Jiang, H., Bodrossy, L., Zhao, J., Murrell, J.C., Xing, X.-H., 2009. Diversity and activity of methanotrophs in alkaline soil from a Chinese coal mine. FEMS Microbiology Ecology 70, 196–207.
Henckel, T., Jäckel, U., Schnell, S., Conrad, R., Jackel, U., Schnell, S., Conrad, R., Jäckel, U., Schnell, S., Conrad, R., 2000. Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. Applied and Environmental Microbiology 66, 1801–1808. https://doi.org/10.1128/AEM.66.5.1801-1808.2000.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One. https://doi.org/10.1371/journal.pone.0169748.
Hermans, S.M., Buckley, H.L., Case, B.S., Curran-cournane, F., Taylor, M., 2017. Bacteria as Emerging Indicators of Soil Condition 83, 1–13.
Ho, A., Kerckhof, F.M., Luke, C., Reim, A., Krause, S., Boon, N., Bodelier, P.L.E., 2013. Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. Environmental Microbiology Reports 5, 335–345. https://doi.org/10.1111/j.1758-2229.2012.00370.x.
Horner-Devine, M.C., Lage, M., Hughes, J.B., Bohannan, B.J.M., 2004. A taxa-area relationship for bacteria. Nature 432, 750–753.
Hu, H.-W.H.W., Chen, D., He, J.Z.J.-Z., 2015. Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. FEMS Microbiology Reviews 39, 729–749. https://doi.org/10.1093/femsre/fuv021.
Jenny, H., 1994. Factors of Soil Formation: a System of Quantitative Pedology. Courier Corporation.
Judd, C.R., Koyama, A., Simmons, M.P., Brewer, P., von Fischer, J.C., 2016. Co-variation in methanotroph community composition and activity in three temperate grassland soils. Soil Biology and Biochemistry 95, 78–86. https://doi.org/10.1016/j.soilbio.2015.12.014.
Karunaratne, S.B., Bishop, T.F.A., Baldock, J.A., Odeh, I.O.A., 2014. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219, 14–23. https://doi.org/10.1016/j.geoderma.2013.12.005.
Kim, H.J., Galeva, N., Larive, C.K., Alterman, M., Graham, D.W., 2005. Purification and physical-chemical properties of methanobactin: a chalkophore from Methylosinus trichosporium OB3b. Biochemistry 44, 5140–5148.
Knief, C., 2015. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. Frontiers in Microbiology 6. https://doi.org/10.3389/fmicb.2015.01346.
Knittel, K., Boetius, A., 2009. Anaerobic oxidation of methane: progress with an unknown process. Annual Review of Microbiology 63, 311–334.
Kolb, S., 2009. The quest for atmospheric methane oxidizers in forest soils. Environmental Microbiology Reports 1, 336–346. https://doi.org/10.1111/j.1758-2229.2009.00047.x.
Kolb, S., Horn, M.A., 2012. Microbial CH4 and N2O consumption in acidic wetlands. Frontiers in Microbiology 3, 78.
Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S., Grossart, H.-P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Front. Microbiol. 5.
Krause, S., Lüke, C., Frenzel, P., 2009. Spatial heterogeneity of methanotrophs: a geostatistical analysis of pmoA-based T-RFLP patterns in a paddy soil. Environmental Microbiology Reports 1, 393–397.
Krause, S., Meima-Franke, M., Hefting, M.M., Bodelier, P.L.E., 2013. Spatial patterns of methanotrophic communities along a hydrological gradient in a riparian wetland. FEMS Microbiology Ecology 86, 59–70. https://doi.org/10.1111/1574-6941.12091.
Kurm, V., Van Der Putten, W.H., De Boer, W., Naus-Wiezer, S., Gera Hol, W.H., 2017. Low abundant soil bacteria can be metabolically versatile and fast growing. Ecology 98, 555–564. https://doi.org/10.1002/ecy.1670.
Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. European Journal of Soil Science 57, 787–799. https://doi.org/10.1111/j.1365-2389.2005.00768.x.
Larsen, P.E., Field, D., Gilbert, J.A., 2012. Predicting bacterial community assemblages using an artificial neural network approach. Nature Methods 9, 621–625. https://doi.org/10.1038/nmeth.1975.
Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271–280. https://doi.org/10.1007/s004420100716.
Lennon, J.T., Aanderud, Z.T., Lehmkuhl, B.K., Schoolmaster, D.R., 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93, 1867–1879. https://doi.org/10.1890/11-1745.1.
Li, C., Narayanan, V., Harriss, R.C., 1996. Model estimates of nitrous oxide emissions from agricultural lands in the United States. Global Biogeochemical Cycles 10, 297–306.
Lieberman, R.L., Rosenzweig, A.C., 2004. Biological methane oxidation: regulation, biochemistry, and active site structure of particulate methane monooxygenase. Critical Reviews in Biochemistry and Molecular Biology 39, 147–164.
Lilly, A., Bell, J., Hudson, G., Nolan, A., Towers, W., 2010a. National Soil Inventory of Scotland 1 (NSIS_1): Site Location, Sampling and Profile Description Protocols (1978-1988). Tecnical Bulletin. Macaulay Institute.
Lilly, A., Hudson, G., Hough, R.L., 2010b. Developing methods to detect change in the soil resource of a country with a Northern, Temperate Boreal Climate (Scotland). In: Proceedings of the 19th World Congress of Soil Science: Soil Solutions for a Changing World, Brisbane, Australia, 1-6 August 2010. Symposium 1.5. 1 Quantitative Monitoring of Soil Change. International Union of Soil Sciences (IUSS), c/o Institut für Bodenforschung, Universität für Bodenkultur, pp. 43–46.
Lima, A.B., Muniz, A.W., Dumont, M.G., 2014. Activity and abundance of methane-oxidizing bacteria in secondary forest and manioc plantations of Amazonian Dark Earth and their adjacent soils. Frontiers in Microbiology 5.
Lüke, C., Krause, S., Cavigiolo, S., Greppi, D., Lupotto, E., Frenzel, P., 2010. Biogeography of wetland rice methanotrophs. Environmental Microbiology 12, 862–872. https://doi.org/10.1111/j.1462-2920.2009.02131.x.
Martineau, C., Pan, Y., Bodrossy, L., Yergeau, E., Whyte, L.G., Greer, C.W., 2014. Atmospheric methane oxidizers are present and active in Canadian high Arctic soils. FEMS Microbiology Ecology 89, 257–269.
Narihiro, T., Hori, T., Nagata, O., Hoshino, T., Yumoto, I., Kamagata, Y., 2011. The impact of aridification and vegetation type on changes in the community structure of methane-cycling microorganisms in Japanese wetland soils. Bioscience Biotechnology and Biochemistry 75, 1727–1734.
Nazaries, L., 2011. Impact of Land-use Changes on the Methanotrophic Community Structure. University of Warwick. University of Warwick.
Nazaries, L., Murrell, J.C., Millard, P., Baggs, L., Singh, B.K., 2013a. Methane, microbes and models: fundamental understanding of the soil methane cycle for future predictions. Environmental Microbiology 15, 2395–2417. https://doi.org/10.1111/1462-2920.12149.
Nazaries, L., Pan, Y., Bodrossy, L., Baggs, E.M., Millard, P., Murrell, J.C., Singh, B.K., 2013b. Evidence of microbial regulation of biogeochemical cycles from a study on methane flux and land use change. Applied and Environmental Microbiology 79, 4031–4040. https://doi.org/10.1128/AEM.00095-13.
Nazaries, L., Tate, K.R., Ross, D.J., Singh, J., Dando, J., Saggar, S., Baggs, E.M., Millard, P., Murrell, J.C., Singh, B.K., 2011. Response of methanotrophic communities to afforestation and reforestation in New Zealand. The ISME Journal 5, 1832–1836. https://doi.org/10.1038/ismej.2011.62.
Nazaries, L., Tottey, W., Robinson, L., Khachane, A., Al-Soud, W.A., Sørenson, S., Singh, B.K., 2015. Shifts in the microbial community structure explain the response of soil respiration to land-use change but not to climate warming. Soil Biology and Biochemistry 89, 123–134. https://doi.org/10.1016/j.soilbio.2015.06.027.
Nelson, M.A., Bishop, T.F.A., Triantafilis, J., Odeh, I.O.A., 2011. An error budget for different sources of error in digital soil mapping. European Journal of Soil Science 62, 417–430. https://doi.org/10.1111/j.1365-2389.2011.01365.x.
Nielsen, A.K., Gerdes, K., Murrell, J.C., 1997. Copper-dependent reciprocal transcriptional regulation of methane monooxygenase genes in Methylococcus capsulatus and Methylosinus trichosporium. Molecular Microbiology 25, 399–409. https://doi.org/10.1046/j.1365-2958.1997.4801846.x.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Package "vegan." Community Ecology Package 2.
Orton, T.G., Pringle, M.J., Page, K.L., Dalal, R.C., Bishop, T.F.A., 2014. Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. Geoderma 230–231, 119–130. https://doi.org/10.1016/j.geoderma.2014.04.016.
Powell, J.R., Karunaratne, S., Campbell, C.D., Yao, H., Robinson, L., Singh, B.K., 2015a. Deterministic processes vary during community assembly for ecologically dissimilar taxa. Nature Communications 6, 8444. https://doi.org/10.1038/ncomms9444.
Powell, J.R., Welsh, A., Hallin, S., Allison, S.D., 2015b. Microbial functional diversity enhances predictive models linking environmental parameters to ecosystem properties. Ecology 96https://doi.org/10.1890/14-1127.1. 150107134727006.
Pratscher, J., Vollmers, J., Wiegand, S., Dumont, M.G., Kaster, A., 2018. Unravelling the identity, metabolic potential and global biogeography of the atmospheric methane-oxidizing upland soil cluster α. Environmental Microbiology. https://doi.org/10.1111/1462-2920.14036.
Ramette, A., 2009. Quantitative community fingerprinting methods for estimating the abundance of operational taxonomic units in natural microbial communities. Applied and Environmental Microbiology 75, 2495–2505. https://doi.org/10.1128/AEM.02409-08.
Ramette, A., Tiedje, J.M., 2007. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. Microbial Ecology 53, 197–207. https://doi.org/10.1007/s00248-005-5010-2.
Ranjard, L., Dequiedt, S., Jolivet, C., Saby, N.P.A., Thioulouse, J., Harmand, J., Loisel, P., Rapaport, A., Fall, S., Simonet, P., Joffre, R., Bouré, N.C.-P.P., Maron, P.-A.A., Mougel, C., Martin, M.P., Toutain, B., Arrouays, D., Lemanceau, P., 2010. Biogeography of soil microbial communities: a review and a description of the

ongoing French national initiative. Agronomy for Sustainable Development 30, 359–365. https://doi.org/10.1007/978-94-007-0394-0_37.
Reay, D.S., Smith, P., van Amstel, A., 2010. Methane and Climate Change. Earthscan.
Saggar, S., Hedley, C.B., Giltrap, D.L., Lambie, S.M., 2007. Measured and modelled estimates of nitrous oxide emission and methane consumption from a sheep-grazed pasture. Agriculture, Ecosystems & Environment 122, 357–365.
Semrau, J.D., DiSpirito, A.A., Gu, W., Yoon, S., 2018. Metals and Methanotrophy. Applied and Environmental Microbiology AEM.02289-17. https://doi.org/10.1128/AEM.02289-17.
Semrau, J.D., Dispirito, A.A., Yoon, S., 2010. Methanotrophs and copper. FEMS Microbiology Reviews 34, 496–531.
Shipley, B., 2016. Cause and Correlation in Biology: a User's Guide to Path Analysis, Structural Equations and Causal Inference with R. Cambridge University Press.
Singh, B.K., Bardgett, R.D., Smith, P., Reay, D.S., 2010. Microorganisms and climate change: terrestrial feedbacks and mitigation options. Nature Reviews Microbiology 8, 779–790.
Singh, B.K., Tate, K.R., Kolipaka, G., Hedley, C.B., Macdonald, C.A., Millard, P., Murrell, J.C., 2007. Effect of afforestation and reforestation of pastures on the activity and population dynamics of methanotrophic bacteria. Applied and Environmental Microbiology 73, 5153–5161.
Singh, B.K., Tate, K.R., Ross, D.J., Singh, J., Dando, J., Thomas, N., Millard, P., Murrell, J.C., 2009. Soil methane oxidation and methanotroph responses to afforestation of pastures with Pinus radiata stands. Soil Biology and Biochemistry 41, 2196–2205.
Siqueira, T., Bini, L.M., Roque, F.O., Marques Couceiro, S.R., Trivinho-Strixino, S., Cottenie, K., 2012. Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. Ecography 35, 183–192. https://doi.org/10.1111/j.1600-0587.2011.06875.x.
Tate, K.R., 2015. Soil methane oxidation and land-use change - from process to mitigation. Soil Biology and Biochemistry 80, 260–272. https://doi.org/10.1016/j.soilbio.2014.10.010.
Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and geography of soil fungi. Science 346.
Thauer, R.K., Shima, S., 2008. Methane as fuel for anaerobic microorganisms. Annals of the New York Academy of Sciences 1125, 158–170.
Thies, J.E., 2007. Soil microbial community analysis using terminal restriction fragment length polymorphisms. Soil Science Society of America Journal 71, 579–591.
Trotsenko, Y.A., Murrell, J.C., 2008. Metabolic aspects of aerobic obligate methanotrophy. Advances in Applied Microbiology 63, 183–229.
Wen, X., Yang, S., Liebner, S., 2016. Evaluation and update of cutoff values for methanotrophic pmoA gene sequences. Archives of Microbiology 198, 629–636.
Whalen, S.C., 2005. Biogeochemistry of methane exchange between natural wetlands and the atmosphere. Environmental Engineering Science 22, 73–94.
Wu, Z., Shi, Y., Zeng, J., Zhang, T., Mu, W., Deng, L., Zhang, Q., Lou, K., 2013. Pattern of soil methanotrophs on the natural vertical zones of the North Tianshan Mountain. Acta Ecologica Sinica 33, 80–86. https://doi.org/10.1016/j.chnaes.2013.01.003.
Yao, H., Campbell, C.D., Chapman, S.J., Freitag, T.E., Nicol, G.W., Singh, B.K., 2013. Multi-factorial drivers of ammonia oxidizer communities: evidence from a national soil survey. Environmental Microbiology 15, 2545–2556. https://doi.org/10.1111/1462-2920.12141.